METHODS FOR COMPUTATION AND ANALYSIS OF MARKOVIAN DYNAMICS ON COMPLEX NETWORKS

by

W. Garrett Jenkinson

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

October, 2013

Abstract

A problem central to many scientific and engineering disciplines is how to deal with noisy dynamic processes that take place on networks. Examples include the ebb and flow of biochemical concentrations within cells, the firing patterns of neurons in the brain, and the spread of disease on social networks. In this thesis, we present a general formalism capable of representing many such problems by means of a master equation. Our study begins by synthesizing the literature to provide a toolkit of known mathematical and computational analysis techniques for dealing with this equation. Subsequently a novel exact numerical solution technique is developed, which can be orders of magnitude faster than the state-of-the-art numerical solver. However, numerical solutions are only applicable to a small subset of processes on networks. Thus, many approximate solution techniques exist in the literature to deal with this problem. Unfortunately, no practical tools exist to quantitatively evaluate the quality of an approximate solution in a given system. Therefore, a statistical tool that is capable of evaluating any analytical or Monte Carlo based approximation to the master equation is developed herein. Finally, we note that larger networks

ABSTRACT

with more complex dynamical phenomena suffer from the same curse of dimensionality as the classical mechanics of a gas. We therefore propose that thermodynamic analysis techniques, adapted from statistical mechanics, may provide a new way forward in analyzing such systems. The investigation focuses on a behavior known as avalanching—complex bursting patterns with fractal properties. By developing thermodynamic analysis techniques along with a potential energy landscape perspective, we are able to demonstrate that increasing intrinsic noise causes a phase transition that results in avalanching. This novel result is utilized to characterize avalanching in an epidemiological model for the first time and to explain avalanching in biological neural networks, in which the cause has been falsely attributed to specific neural architectures. This thesis contributes to the existing literature by providing a novel solution technique, enhances existing and future literature by providing a general method for statistical evaluation of approximative solution techniques, and paves the way towards a promising approach to the thermodynamic analysis of large complex processes on networks.

Primary Reader: John Goutsias

Secondary Reader: Andrew Feinberg

Acknowledgments

Ostensibly, this doctoral thesis is the product of six years of education, excogitation, and effort; however the reality is that these six years of work at Johns Hopkins University were propped up by the preceding twenty-two years of learning. Likewise, although this thesis bears my name, none of it would have been possible without an amazing complex network of support. Complete enumeration of this network is beyond my space limitations, so I will attempt to discuss its hubs, which have played the most critical role.

The most supportive member of my network is my fiancée, Dr. Lara Walkoff, who has been by my side for the past decade to see me grow from a high school math geek to the absent-minded academic of the present. She makes my journey through life enjoyable and exciting, while her brilliance and work ethic set the precedent that I continually strive to achieve.

By far, the most important professional node in my network is my advisor, Prof. John Goutsias. He considered me a peer from day one, and our often contentious and heated debates about science (which were never taken personally) served to pu-

rify our scientific claims like gold from fire. Despite our enormous mutual respect, neither of us has trusted a scientific claim from the other until we have understood it individually. Peer review started long before we submitted our papers—and this, in my opinion, is how science should always be conducted. I will be forever grateful that John has fostered my mathematical thinking and skills to a level that I never imagined possible. Graduate school is intended to be a modern apprenticeship, and I certainly feel that John has succeeded in bringing our minds to a convergent state. More often our thoughts are in lockstep, and thus our conversations can occur in rapid shorthand that would surely be incomprehensible to an outside observer. His mentorship has been invaluable, and I cherish the fact that we will undoubtedly remain life-long friends.

I am indebted to my second reader, Prof. Andy Feinberg, whose ebullient enthusiasm demonstrates the true meaning of life-long learning, even after a lifetime of astonishing accomplishments. I look forward to many years of collaboration where our productivity will only be matched by the enjoyment of the voyage.

I would also like to express gratitude to my committee members, Prof. Howard Wienert and Prof. Pablo Iglesias, for taking the time to improve and evaluate this work.

Graduate school has been made possible by the following funding sources: the Wolman family for an Abel Wolman Fellowship, the Institute for Nanobiotechnology for a National Science Foundation (NSF) IGERT Fellowship, the Department

of Defense for an NDSEG Fellowship, Prof. Goutsias for funding me through two NSF Grants, CCF-0849907 and CCF-1217213, and finally the Siebel Foundation for their prestigious Siebel Scholarship.

I would like to extend special thanks to some of the most influential mentors who helped me crystalize a love for teaching. I am grateful to Prof. Carey Priebe at Johns Hopkins, the most talented teacher I have encountered in 23 years of classes, for providing me with a standard to emulate. Prof. Jelena Kovačević and Prof. José Moura at Carnegie Mellon University instilled in me the joy of research as an undergrad. Then, there were those teachers who inspired me early on: Sue Frennesson showed me the joys of mathematics in middle school, Dr. Michael Greene gave me the love of physics in high school, and Marcia Mett brought chemistry to life in the high school laboratory.

Most importantly, my friends and family have made the journey entertaining. The support of my grandparents Ashford and Carolyn Jenkinson has been invaluable, and I always look forward to calling them to share the joy of my latest achievement. Special thanks to my brother and best friend, Gavin, for relocating to Baltimore and keeping the sometimes mundane life of a graduate student amusing.

Dedication

This thesis is dedicated to my parents who, through their emphasis on quality education, infected me with an incurable intellectual bug that prepared me for this career choice. They have taught by example the meaning of hard work and taking pride in one's vocation. This lesson is the main reason I was able to persevere and become the first Dr. Jenkinson in our family. Undoubtedly, the fact that I had an outstanding childhood has also contributed to my sanity, which most people would have lost from spending so many hours toiling upwards into the night. Their continuing support, encouragement and interest in my arcane pursuits make me joyous beyond words and are causal features of my perpetual smile. Alan and Debbie, I love you both, and hope you enjoy reading my monolithic tribute to the greatest parents in the world—although I will understand if you choose to skim some of the details!

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

We live in a networked world. We are surrounded by large-scale networks that wield enormous influence over our daily lives; we travel on infrastructure networks, communicate by telecom networks or the Internet, are governed by the political will of a complex social network, and become infected by diseases that spread over the network of people who come into physical or sexual contact. We are made of networks; our tissues are perfused by vast supply and drainage networks of the circulation and lymph systems, and our cells operate through chemical reaction networks that are so complex they can create an entire human being from a single cell. Our cognition and ability to understand networks arises from a network; the network of firing neurons in our brain is capable of computations beyond that of our most powerful supercomputers.

In this thesis, we study processes on networks—quantities that change over time in a way that is constrained by the network on which it "lives." An example would be a disease that spreads over a social network. The process here is the categorization of every person on the social network with respect to a disease: they are either infected with strep throat or they are healthy and able to catch strep throat. This process changes with time—we might be healthy today, and sick tomorrow.

The network constrains this process—if we are in Baltimore, we cannot become infected by our friends who are sick in Boston. Therefore, although the networks are complex structures, they actually serve to simplify dynamical processes by constraining the number of possible changes that can occur at a given time.

In addition to being constrained by an underlying network, the processes that we study are intrinsically noisy. They are stochastic and cannot be predicted with absolute certainty—when our sick friends sneeze on us it does not mean we are guaranteed to become sick. The best we can do is to quantify our certainty with probabilities. Engineers might be tempted to view such noisy processes as problematic or dysfunctional, since we attempt to design man-made systems to mitigate the effects of noise. To most engineers, noise is a bad thing. In this thesis, one observation we make is that nature disagrees with this perspective; instead it has evolved our brains to utilize the intrinsic noisiness of our neurons to create beautiful, fractal bursting patterns of activity that are thought to play a crucial role in information processing, storage, and learning. These complex dynamics are simply not possible in a noise-free environment (see Chapter 5).

As with any rigorous probe into the world around us, our study takes place in the language of mathematics. Specifically, we study the previous stochastic processes through the mathematical framework of Markov processes, which has been extensively investigated for more than a century. The equation of interest here is the forward Kolmogorov equation. What makes our study modern and interesting is the fact that we focus on processes that are nonlinear and constrained by an underlying network topology. In particular, chemical physicists have studied a special case of the forward Kolmogorov equation, known as the chemical master equation, which is capable of describing nonlinear chemical reactions taking place on large networks [1]. In this work, we present the coherent framework of *Markovian reaction networks* that is capable of representing stochastic processes on networks using a master equation.

The study of the master equation is an area of research that has been evolving for a long time. Its applicability to many scientific and engineering disciplines has led to parallel and

often independent developments, which have recently reached a critical mass due to unprecedented advancements in modern experimental procedures and computational capabilities. By presenting a rigorous mathematical framework, developments across many fields may be brought together into a single coherent toolkit. Additional developments presented in this thesis (or in the future) may serve to simultaneously advance many scientific disciplines. Here, we will focus on computationally solving the master equation or analyzing the resulting stochastic processes by means of powerful statistical thermodynamic tools when the solution to the full master equation is intractable.

1.2 Scope and organization of thesis

The tremendous flexibility and generality of Markovian reaction networks make them an excellent mathematical framework for studying stochastic processes on complex networks. The coherency of a single framework means that tools and discoveries made in one field may be readily ported to distant applications. In Chapter 2, we present this rigorous framework, while synthesizing and reviewing the fractionated literature that deals with problems in this framework.

The literature has shown [2] that the generality of Markovian reaction networks allows even networks limited to simple components (i.e., mass action propensity functions, see Section 2.2.1) to perform Turing universal computations with arbitrarily small error that becomes zero at the limit of infinite system size [3]. This strength also turns out to be one of the most profound weaknesses of Markovian reaction networks: *there will be no single analytical or even computational method capable of calculating the exact solution of the underlying master equation in complete generality using finite resources.* As a consequence, the development of accurate and computationally feasible techniques for studying the dynamic behavior of large nonlinear Markovian reaction networks is still the most important and challenging problem in this field of research. Therefore, while reading this thesis, one must keep in mind three crucial points:

1. There will never be a "silver bullet" algorithm which can efficiently solve every master equation,

so we do not seek one in this work.

2. This thesis does not exist in a bubble, and thus the rich existing literature should be utilized and enhanced by this work.

3. The full solution to the master equation may not be the most useful tool in the analysis of these stochastic systems, so we will attempt to develop thermodynamic tools that will prove useful in these instances.

The way to deal with Point 1 is to focus on specialized structures that may be present in many reaction networks of interest. By exploiting these structures, we are able to develop rigorous solution techniques tailored to the specific application at hand. In Chapter 3, we develop such an approach, which we refer to as the implicit Euler (IE) method. This method numerically calculates the exact solution to the master equation, up to a desired precision, by exploiting the structure of the master equation that governs a more informative stochastic process, which has been largely overlooked in the literature. More specifically, we discuss in Section 2.3.2 how the master equation on a finite state space can be viewed as a large system of sparse, linear ordinary differential equations (ODEs). The most successful numerical tool in the literature exploits the linearity and the sparseness of the problem, resulting in an algorithm that is preferable to other solution methods whenever the state space is small enough.

In Chapter 3, we observe that the usual Markov process examined in the literature is actually less informative than a related counting process that has more inherent structure. Specifically, the probability mass moves monotonically through the state space of a counting process, never moving backwards towards the origin. We can exploit this by appropriately ordering the state space (e.g., lexicographically), which adds further structure to the linear system of ODEs. The additional structure is the fact that the sparse matrix is now also triangular. Therefore, the (usually computationally expensive) IE method of solving linear ordinary differential equations becomes extremely efficient, since inversion of the generator matrix (the bottleneck of the IE method) can be calculated

efficiently via recursive back-substitution. Therefore, the IE method can enjoy a speedup that is orders of magnitude beyond the state-of-the-art algorithm, while also experiencing many additional gains, such as numerical stability while ensuring that the solution of the master equation remains a probability distribution (which is a significant problem in the state-of-the-art algorithm). However, this added speed and stability is not without limitations. The applicability of the IE method depends on the underlying structure of the reaction network. If the network structure does not satisfy some well-defined criteria, the IE method should be avoided due to the explosion in the size of the state space of the counting process. We demonstrate, however, that for a wide class of systems in epidemiology, the IE method will be superior to the state-of-the-art algorithm.

The are two ways to deal with Point 2. First, as mentioned earlier, we use Chapter 2 to bring together a vast and often non-communicating literature, allowing the open problems in the field to be more readily identified. One glaring problem we identified is that the large number of approximate solution techniques are often justified by theoretical limit results which provide little guidance to a practitioner with a particular network of interest. The practitioner is often left to blindly use an approximate solution technique, with no way of knowing if egregious errors are accumulating, leading to papers being published with erroneous results. In fact, a published work using a common approximation technique (i.e., the linear noise approximation, see Section 2.3.4) existed in the literature for many years without anyone identifying the large errors caused by the approximation, until we applied our exact solution technique from Chapter 3 to the same problem and identified the errors introduced by the approximation method. This experience motivated our second way to deal with Point 2: we develop a tool in Chapter 4 that provides a statistical algorithm which is capable of evaluating the accuracy of a given solution technique in a particular system of interest. This is the first general tool of its kind that is capable of statistically verifying the accuracy of any approximate solution of the master equation. By utilizing the well-known Kolmogorov-Smirnov test statistic, we are able to produce confidence bands around the marginal cumulative distributions that solve the master equation. If an approximation deviates significantly from these

confidence bands, it should be rejected as an inadequate solution technique; whereas, the confidence bands also quantify the extent to which a decent approximation technique can be trusted. In this way, the technique presented in Chapter 4 serves to enhance the rich body of work laid down by previous research efforts, by allowing a practicing scientist to ignore approximations that provide erroneous results, and to quantitatively understand the extent to which a given approximation can be trusted in their system of interest.

Point 3 is well known in the world of statistical mechanics. Consider a gas within a container. Physicists are clear what the complete description of such a system is (ignoring quantum or relativistic effects): Hamiltonian dynamics. The state of the gas is given by the position and momentum of each of its $\sim 10^{23}$ molecules, which evolve in time according to Newton's laws of motion. For the moment, let us ignore the computational complexity of solving this exceptionally large system of equations. Let us assume we have a cosmic computer which gives us the answer in a reasonable time. A mathematician will be surely pleased, having the solution to such a complex problem; however, as a physicist or engineer trying to understand this system, we will be no more informed than before having the solution. Understanding the dynamics in such a high-dimensional state space is simply too complex. This is where statistical mechanics enters the picture, since a few numbers (calculated through the use of statistical averages), such as temperature, pressure, entropy, and internal energy, can succinctly describe this enormously complex system in a way that is easily interpretable.

Likewise, we are aware in this thesis that—analogously to the Hamiltonian equation for the gas—the master equation and its corresponding solution may not actually be the most practical description of a large and highly complex process taking place on a network. Thus, we present in Chapter 5 a special class of Markovian reaction networks, which we call *leaky Markovian networks*. These networks have binary state variables, such as a neuron that can be active or quiescent. In the brain however, knowing the probability for every neuron being active or quiescent would be akin to knowing the position and momentum of every molecule of gas—information overload, since the

state space grows exponentially as 2^N, where N is the number of nodes in our network. Therefore, we develop in Chapter 5 statistical mechanical tools for analyzing leaky Markovian networks and demonstrate that intrinsic noise plays a fundamental and previously unknown role in these systems.

In particular, we demonstrate in Chapter 5 that intrinsic noise induces a phase transition in leaky Markovian networks leading to avalanching—a complex bursting pattern with fractal properties. By using the notion of a potential energy landscape, we demonstrate that noise is not only capable of producing uphill movements on the potential energy landscape, but it is also capable of warping the landscape itself. In leaky Markovian networks, this warping alters the global stability properties of the network, leading to a phase transition that results in avalanching being the predominant system behavior. As examples of leaky Markovian networks, we study the spread of Methicillin-resistant Staphylococcus aureus (MRSA) infections and the activity patterns of biological neural networks. The first example provides the first characterization of avalanching in epidemiological networks. On the other hand, the second example sheds light on the cause of avalanching in neural networks where the importance of avalanches is an open field of research, but their presence in real data is often observed [4].

We feel that statistical mechanics tools applied to these networks will pave the way towards a new paradigm in the study of processes on large networks, where the intractable solution of the master equation is no longer the central focus of research. Thermodynamic quantities, such as the potential energy landscape, entropy, internal energy, pressure, and bulk modulus, will become a primary focus of algorithm development. For example, the original master equation for the spread of MRSA in a population of 300 individuals is hampered by the fact that the time-evolving probability distribution "lives" in a state space with 2^{300} elements. Therefore, if a hypothetical cosmic computer were to provide the stationary solution to the master equation, one would need to store (and then analyze) $2^{300} \simeq 10^{90}$ probability values. Note that there are only $\sim 10^{80}$ molecules in the known universe, so even the cosmic computer is not capable of storing this result. Instead, after thermodynamic coarse graining of the system, we show how it can be analyzed by (among other

quantities) an easily computable potential energy landscape that lives in a state space with only 301 elements. We therefore realize in Chapter 5 that the answer most people would think they are seeking (i.e., the full solution to the original master equation) is not the answer that most people would want to receive. By shifting focus and effort towards computing thermodynamic summaries of system properties, we can find greater clarity at reduced cost.

In Chapter 6, we finally provide concluding remarks and discuss future work in the area of Markov processes on complex networks. For clarity of presentation, we relegate the more involved mathematical or computational details throughout this thesis to the Appendices.

Chapter 2

Markovian Reaction Networks: A Coherent Framework

In this chapter, we provide necessary background on Markovian reaction networks and the master equation formalism that we will be using throughout this thesis[1]. The mathematical and computational framework of Markovian reaction networks encompasses problems at the cutting edge of a diverse number of scientific fields such as: biochemistry, pharmacokinetics, epidemiology, ecology, sociology, neurobiology, multi-agent networks, and evolutionary game theory; a more sweeping review of this framework may be found in [5].

Our main goal is to provide a comprehensive and coherent coverage of recently developed approaches and methods to model complex nonlinear Markovian reaction networks and analyze their dynamic behavior. To achieve this, we first review in Section 2.1 a general framework for modeling Markovian reaction networks and subsequently discuss specific examples within this framework in Section 2.2. In Section 2.3, we provide a review of the relevant numerical and computational techniques available for estimating or approximating the solution of the master equation. In addition,

[1]Materials in this chapter are reprinted from "Markovian dynamics on complex reaction networks", volume 529, issue 2, by John Goutsias and Garrett Jenkinson, *Physics Reports*, pp. 199-264, Copyright (2013), with permission from Elsevier.

we review in Section 2.4 several mathematical facts pertaining to the mesoscopic (probabilistic) behavior of the master equation. These facts are well-known from the theory of Markov processes, but we recast them here in the more specific form dictated by the framework of Markovian reaction networks. In Section 2.5, we discuss a recently developed approach for studying the stationary behavior of Markovian reaction networks using a potential energy landscape perspective, whereas we present in Section 2.6 an introduction to the emerging theory of thermodynamic analysis of Markovian reaction networks.

2.1 Reaction networks

2.1.1 Chemical systems and reaction networks

Networks of chemical reactions are used extensively to model biochemical activity in cells. It turns out that many physical and man-made systems of interest to science and engineering can be viewed as special cases of chemical reaction networks when it comes to mathematical and computational analysis. For this reason, chemical reaction networks can serve as archetypal systems when studying dynamics on complex networks.

A chemical reaction system is comprised of a (usually) large number of molecular species and chemical reactions. A group of molecular species, known as *reactants*, interact through a chemical reaction to create a new set of molecular species, known as *products*. In general, we can think of a set of chemical reactions as a system that consists of N molecular species $X_1, X_2, \ldots, X_N$ that interact through M coupled reactions of the form:

$$\sum_{n\in\mathcal{N}} \nu_{nm} X_n \;\rightarrow\; \sum_{n\in\mathcal{N}} \nu'_{nm} X_n, \quad m \in \mathcal{M}, \tag{2.1.1}$$

where $\mathcal{N} := \{1, 2, \ldots, N\}$ and $\mathcal{M} := \{1, 2, \ldots, M\}$. The quantities $\nu_{nm} \geq 0$ and $\nu'_{nm} \geq 0$ are known as the *stoichiometric coefficients* of the reactants and products, respectively. These coefficients tell us how many molecules of the n-th species are consumed or produced by the m-th reaction. In

particular, the notation used in Eq. (2.1.1) implies that occurrence of the m-th reaction changes the molecular count of species X_n by $s_{nm} := \nu'_{nm} - \nu_{nm}$, where s_{nm} is known as the *net* stoichiometric coefficient.

The inter-connectivity between components in a chemical reaction system can be graphically represented as a network [6,7] and, more specifically, by means of a directed, weighted, bipartite graph. Since molecular species react with each other to produce other molecular species, we can refer to this network in more general terms as a *reaction network*.

To illustrate how we can map a chemical reaction system to a network, let us consider the following reactions that correspond to a quadratic autocatalator with positive feedback [8]:

$$\begin{aligned} \mathrm{S} &\rightarrow \mathrm{P} \\ \mathrm{D} + \mathrm{P} &\rightarrow \mathrm{D} + 2\mathrm{P} \\ 2\mathrm{P} &\rightarrow \mathrm{P} + \mathrm{Q} \\ \mathrm{P} + \mathrm{Q} &\rightarrow 2\mathrm{Q} \\ \mathrm{P} &\rightarrow \emptyset \\ \mathrm{Q} &\rightarrow \emptyset, \end{aligned} \tag{2.1.2}$$

where the last two reactions indicate the degradation of molecules P and Q. This chemical reaction system is comprised of $N = 4$ molecular species and $M = 6$ reactions. We can (arbitrarily) label the molecular species as $X_1 = \mathrm{S}$, $X_2 = \mathrm{P}$, $X_3 = \mathrm{D}$, $X_4 = \mathrm{Q}$, and the reactions as $1, 2, \ldots, 6$. We can now represent the system by the network of interactions depicted in Fig. 2.1. This network consists of two types of nodes: those representing the molecular species (white circles) and those representing the reactions (black circles). The directed edges represent interactions between molecular species and reactions and, naturally, connect only white nodes with black nodes. Edges emanating from white nodes and incident to black nodes correspond to the reactants associated with a particular reaction, whereas, edges emanating from black nodes and incident to white nodes correspond to the products of that reaction. Edges are labeled by their weights, which correspond to the stoichiometric coefficients associated with the molecular species represented by the white nodes and the reactions

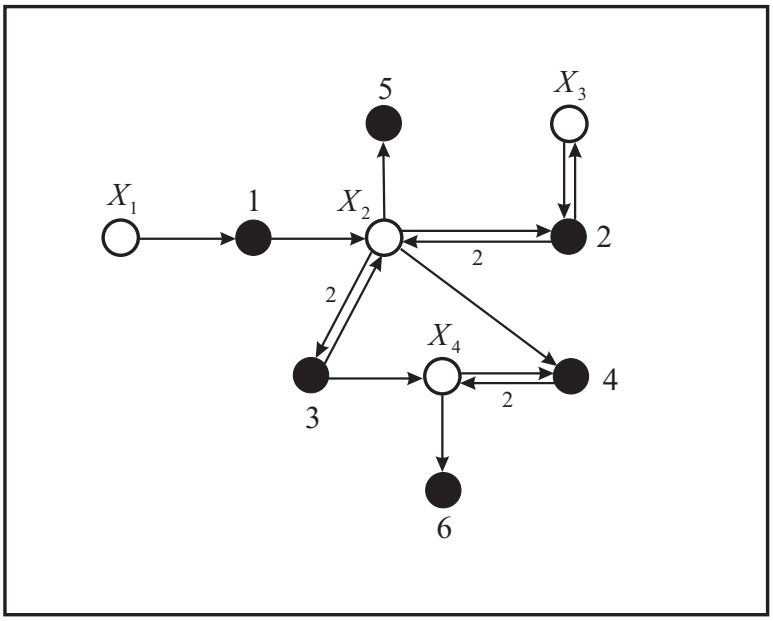

Figure 2.1: A directed, weighted, bipartite graphical representation of the chemical reaction system given by Eq. (2.1.2). The molecular species are represented by the white nodes, whereas, the reactions are represented by the black nodes. Edges emanating from white nodes and incident to black nodes correspond to the reactants associated with a particular reaction, whereas, edges emanating from black nodes and incident to white nodes correspond to the products of that reaction.

represented by the corresponding black nodes. For simplicity, an edge is not labeled when the value of the associated stoichiometric coefficient is one.

An alternative representation of a reaction network is by means of the two $N \times M$ *stoichiometric matrices* $\mathbb{V}$ and $\mathbb{V}'$ with elements ν_{nm} and ν'_{nm}, respectively. These matrices play a similar role as the adjacency matrix of a simple graph [7]. For the reaction network depicted in Fig. 2.1, we have that

$$\mathbb{V} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbb{V}' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \end{bmatrix}. \tag{2.1.3}$$

It is not difficult to see that, given the two stoichiometric matrices $\mathbb{V}$ and $\mathbb{V}'$, we can uniquely construct the chemical reaction system given by Eq. (2.1.2) and, therefore, the network depicted in Fig. 2.1. Hence, knowledge of the two stoichiometric matrices completely specifies the network

topology. Note that a quick glance of these matrices may allow us to make some interesting observations about the chemical reaction system at hand. For example, the fact that all but one of the elements of the first row of matrix $\mathbb{V}$ are zero indicates that the molecular species X_1 is a reactant only in one reaction, whereas, the fact that the first row of matrix $\mathbb{V}'$ is zero indicates that this species is not produced by any reaction. Moreover, the last two zero columns of matrix $\mathbb{V}'$ indicate that reactions 5 and 6 do not result in any products (i.e., they act as sink nodes).

Although the mathematical study of the topological structure of a reaction network is an important topic of research, we will not consider this problem here. Moreover, we will not consider situations in which the topology of the network varies with time. The reader is referred to [7] and the references therein for such topological considerations. Instead, our objective is to discuss mathematical methods and computational techniques for the modeling and analysis of the *dynamic* behavior of reaction networks.

2.1.2 Stochastic dynamics on reaction networks

In many reaction networks of interest, the underlying reactions may occur at random times. If $Z_m(t)$ denotes the number of times that the m-th reaction occurs within the time interval $[0, t)$, then $\{Z_m(t), t \geq 0\}$ will be a random counting process [9]. By convention, we set $Z_m(0) = 0$ (i.e., the reaction never occurs before the initial time $t = 0$). We can employ the $M \times 1$ random vector $\boldsymbol{Z}(t)$ with elements $Z_m(t)$, $m = 1, 2, \ldots, M$, to characterize the state of the system at time $t > 0$. $Z_m(t)$ is usually referred to as the *degree of advancement* (DA) of the m-th reaction [1]. For this reason, we refer to the multivariate counting process $\{\boldsymbol{Z}(t), t > 0\}$ as the *DA process*.

An alternative way to characterize a reaction network is by using the $N \times 1$ random state vector

$$\boldsymbol{X}(t) := \boldsymbol{x}_0 + \mathbb{S}\boldsymbol{Z}(t), \quad t \geq 0, \tag{2.1.4}$$

where $\mathbb{S} := \mathbb{V}' - \mathbb{V}$ is the *net* stoichiometric matrix of the reaction network and $\boldsymbol{x}_0$ is some known value of $\boldsymbol{X}(t)$ at time $t = 0$. Usually, the n-th element $X_n(t)$ of $\boldsymbol{X}(t)$ represents the population

number of the n-th species present in the system at time t, although this may not be true in certain problems. We will be referring to the multivariate stochastic process $\{\boldsymbol{X}(t), t > 0\}$ as the *population process*. For a given initial population vector $\boldsymbol{x}_0$, Eq. (2.1.4) allows us to uniquely determine the random population vector $\boldsymbol{X}(t)$ from the DAs $\boldsymbol{Z}(t)$, provided that $\boldsymbol{Z}(t)$ is almost sure finite.

A large class of reaction networks can be characterized by Markovian dynamics, in which case we refer to them as *Markovian reaction networks*. Markovian reaction networks are based on the fundamental premise that, for a sufficiently small dt, the probability of one reaction to occur within the time interval $[t, t + dt)$ is proportional to dt, with proportionality factor that depends only on the species population present in the system at time t. Specifically, we have that

$$\Pr\big[\text{one reaction } m \text{ occurs within } [t, t + dt) \mid \boldsymbol{X}(t) = \boldsymbol{x}\big] = \pi_m(\boldsymbol{x})dt + o(dt), \tag{2.1.5}$$

for some function $\pi_m(\boldsymbol{x})$ of the population, known as the *propensity function* [10], where $o(dt)$ is a term that goes to zero faster than dt. Under these assumptions, $\{Z_m(t), t > 0\}$ is a (homogeneous) Markovian counting process with intensity $\pi_m(\boldsymbol{X}(t))$. In particular, the probability $p_{\boldsymbol{Z}}(\boldsymbol{z}; t) := \Pr[\boldsymbol{Z}(t) = \boldsymbol{z} \mid \boldsymbol{Z}(0) = \boldsymbol{0}]$ associated with this process satisfies the following partial differential equation (see Appendix A):

$$\frac{\partial p_{\boldsymbol{Z}}(\boldsymbol{z}; t)}{\partial t} = \sum_{m \in \mathcal{M}} \Big\{\alpha_m(\boldsymbol{z} - \mathbf{e}_m)p_{\boldsymbol{Z}}(\boldsymbol{z} - \mathbf{e}_m; t) - \alpha_m(\boldsymbol{z})p_{\boldsymbol{Z}}(\boldsymbol{z}; t)\Big\}, \quad t > 0, \tag{2.1.6}$$

where

$$\alpha_m(\boldsymbol{z}) := \begin{cases} \pi_m(\boldsymbol{x}_0 + \mathbb{S}\boldsymbol{z}), & \text{if } \boldsymbol{z} \geq \boldsymbol{0} \\ 0, & \text{otherwise}, \end{cases} \tag{2.1.7}$$

and $\mathbf{e}_m$ is the m-th column of the $M \times M$ identity matrix [11–13]. This equation is initialized by setting $p_{\boldsymbol{Z}}(\boldsymbol{z}; 0) = \Delta(\boldsymbol{z})$, where $\Delta(\boldsymbol{z})$ is the Kronecker delta function [i.e., $\Delta(\boldsymbol{0}) = 1$ and $\Delta(\boldsymbol{z}) = 0$, if $\boldsymbol{z} \neq \boldsymbol{0}$]. It turns out that the population process $\{\boldsymbol{X}(t), t > 0\}$ is a Markov process as well with probability $p_{\boldsymbol{X}}(\boldsymbol{x}; t) := \Pr[\boldsymbol{X}(t) = \boldsymbol{x} \mid \boldsymbol{X}(0) = \boldsymbol{x}_0]$ that satisfies the following partial differential equation:

$$\frac{\partial p_{\boldsymbol{X}}(\boldsymbol{x}; t)}{\partial t} = \sum_{m \in \mathcal{M}} \Big\{\pi_m(\boldsymbol{x} - \mathbf{s}_m)p_{\boldsymbol{X}}(\boldsymbol{x} - \mathbf{s}_m; t) - \pi_m(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x}; t)\Big\}, \quad t > 0, \tag{2.1.8}$$

initialized by $p_{\boldsymbol{X}}(\boldsymbol{x};0)=\Delta(\boldsymbol{x}-\boldsymbol{x}_0)$, where $\boldsymbol{s}_m$ is the m-th column of the net stoichiometric matrix $\mathbb{S}$. For notational simplicity, we hide the dependency of $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ on $\boldsymbol{x}_0$. Most often, Eq. (2.1.6) and Eq. (2.1.8) are referred to as *master equations* although they are both special cases of a differential form of the Chapman-Kolmogorov equations in the theory of Markov processes (see Appendix A). Note that the solution $q_{\boldsymbol{X}}(\boldsymbol{x};t)$ of Eq. (2.1.8), initialized with an arbitrary probability mass function $q(\boldsymbol{x})$, is related to the solution $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{x}_0,t)$ of Eq. (2.1.8), initialized with $\Delta(\boldsymbol{x}-\boldsymbol{x}_0)$, by $q_{\boldsymbol{X}}(\boldsymbol{x};t)=\sum_{\boldsymbol{x}_0}p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{x}_0,t)q(\boldsymbol{x}_0)$. Therefore, it suffices to only calculate $p_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{x}_0,t)$, for every $\boldsymbol{x}_0$ such that $q(\boldsymbol{x}_0)\neq 0$. For this reason, we focus our discussion on solving Eq. (2.1.8) initialized with $\Delta(\boldsymbol{x}-\boldsymbol{x}_0)$.

The previous master equations provide a suggestive interpretation on how the probabilities $p_{\boldsymbol{Z}}(\boldsymbol{z};t)$ and $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ evolve as a function of time. For example, Eq. (2.1.8) implies that the probability $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ of the population process $\boldsymbol{X}(t)$ taking value $\boldsymbol{x}$ increases during the time interval $[t,t+dt)$ by an amount $dt\sum_{m\in\mathcal{M}}\pi_m(\boldsymbol{x}-\mathbf{s}_m)p_{\boldsymbol{X}}(\boldsymbol{x}-\mathbf{s}_m;t)$ due to possible transitions from states $\boldsymbol{x}-\boldsymbol{s}_m$, $m\in\mathcal{M}$, at time t, to state $\boldsymbol{x}$ at time $t+dt$. However, during the same time period the probability $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ also decreases by an amount $dt\sum_{m\in\mathcal{M}}\pi_m(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)$ due to possible transitions from state $\boldsymbol{x}$ at time t to states $\boldsymbol{x}+\boldsymbol{s}_m$, $m\in\mathcal{M}$, at time $t+dt$. Note finally that, in most practical situations, the elements of $\boldsymbol{x}$ are limited to being inside a finite set (e.g., if x_n counts the number of individuals, then it will be non-negative and bounded from above by the total number of allowed individuals). As a consequence, if an element of $\boldsymbol{x}$ takes value outside the allowable range, then the probability of this state and the propensity to enter this state will both be zero [i.e., $p_{\boldsymbol{X}}(\boldsymbol{x};t)=0$, for all t, and $\pi_m(\boldsymbol{x}-\boldsymbol{s}_m)=0$, for all $m\in\mathcal{M}$].

Although the DA process uniquely determines the population process via Eq. (2.1.4), the opposite is not true in general. This is due to the fact that the matrix $\mathbb{S}^T\mathbb{S}$ may not be invertible. Invertibility of $\mathbb{S}^T\mathbb{S}$ is only possible when the nullity of $\mathbb{S}$ is zero, in which case $\boldsymbol{Z}(t)=(\mathbb{S}^T\mathbb{S})^{-1}\mathbb{S}^T[\boldsymbol{X}(t)-\boldsymbol{x}_0]$ and the DA process can be uniquely determined from the population process. Therefore, we can consider the DA process to be more informative in general than the population process. Note that, if the solution $p_{\boldsymbol{Z}}(\boldsymbol{z};t)$ of the master equation (2.1.6) is known,

then we can calculate the probability mass function $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ without having to solve the master equation (2.1.8). Since we are dealing with discrete random variables, we have that

$$p_{\boldsymbol{X}}(\boldsymbol{x};t) \;=\; \sum_{\boldsymbol{z}\in\mathcal{B}(\boldsymbol{x})} p_{\boldsymbol{Z}}(\boldsymbol{z};t), \quad \text{for } \; t \geq 0, \tag{2.1.9}$$

where $\mathcal{B}(\boldsymbol{x}) := \{\boldsymbol{z}\colon\ \boldsymbol{x} = \boldsymbol{x}_0 + \mathbb{S}\boldsymbol{z}\}$.

2.2 Examples

We now provide a few examples which clearly demonstrate that the previously discussed general framework for reaction networks, based on Eq. (2.1.1), is sufficiently general to characterize Markovian dynamics on many other important networks. Each example is associated with a set of "species" that affect each other's population by interacting through well-defined "reactions." To determine the DA and population dynamics, we only need to specify the mathematical form of the underlying propensity functions – from these, the dynamics follow by solving Eq. (2.1.6) for $p_{\boldsymbol{Z}}(\boldsymbol{z};t)$ or Eq. (2.1.8) for $p_{\boldsymbol{X}}(\boldsymbol{x};t)$. For a more comprehensive list of examples, see [5].

2.2.1 Biochemical networks

When dealing with biochemical reactions, we usually assume that the system is well-stirred and in thermal equilibrium at fixed volume. It can be shown in this case that the probability of a randomly selected combination of reactant molecules at time t to react through the m-th reaction during the infinitesimally small time interval $[t, t+dt)$ is proportional to dt, with a proportionality factor κ_m known as the *specific probability rate constant* of the reaction [14]. As a consequence,

$$\Pr\big[\text{one reaction } m \text{ occurs within } [t,t+dt) \mid \boldsymbol{X}(t) = \boldsymbol{x}\big] = \kappa_m \gamma_m(\boldsymbol{x})dt + o(dt),$$

where $\gamma_m(\boldsymbol{x})$ is the number of distinct subsets of molecules that can form a reaction complex at time t, given by

$$\gamma_m(\boldsymbol{x}) \;=\; \prod_{n\in\mathcal{N}} \binom{x_n}{\nu_{nm}} \;=\; \prod_{n\in\mathcal{N}} [x_n \geq \nu_{nm}]\, \frac{x_n!}{\nu_{nm}!(x_n-\nu_{nm})!}\ , \tag{2.2.1}$$

with $[a_1 \geq a_2]$ being the Iverson bracket (i.e., $[a_1 \geq a_2] = 1$, if $a_1 \geq a_2$, and 0 otherwise). Note that the Iverson bracket guarantees that a reaction will proceed only if all reactants are present in the system. As a consequence, we obtain the following propensity functions:

$$\pi_m(\boldsymbol{x}) = \kappa_m \prod_{n \in \mathcal{N}} \binom{x_n}{\nu_{nm}}, \quad \text{for} \quad m \in \mathcal{M}, \tag{2.2.2}$$

which are said to follow the *mass-action law*. We use the convention $0! = 1$, so $\binom{x_n}{0} = 1$, indicating that the propensity function only depends on the state of the reactants.

We should note here that certain reactions cannot be adequately characterized by propensity functions that follow the mass-action law. For example, let us consider a reaction $X_1 + X_2 \to X_3$ that can occur only when a molecule X_1 is bound by at least one molecule X_2 at two independent binding sites with the same affinity θ. It can be shown (e.g., see [15]) that the fraction of molecules X_1 bound by X_2 is given by $\theta x_1/(1 + \theta x_1)$. This leads to the following *hyperbolic* propensity function for the reaction:

$$\pi(x_1, x_2) = \frac{\kappa \theta x_1 x_2}{1 + \theta x_1}, \tag{2.2.3}$$

where κ is the associated specific probability rate constant. Clearly, the mathematical form of the propensity function of a given reaction depends on the underlying molecular mechanism.

2.2.2 Epidemiological networks

Epidemiological networks study the spread of infectious diseases or agents through a population of individuals. Although numerous publications can be found on the subject, we refer the reader to [7] for an elementary introduction. For a mathematical review of *deterministic* epidemiological models, see [16], whereas, for a *stochastic* modeling approach to epidemiological modeling, see [17].

To illustrate the connection between epidemiological networks and Markovian reaction networks, we consider the simplest and most widely used model, known as the SIR epidemic model. In this model, an individual in a population can be in one of three states with respect to a dis-

ease: *susceptible* (S), *infected* (I), or *resistant* (R). According to this model, there are two types of interactions that an individual may undergo: (a) if a susceptible individual comes into contact with an infectious individual, the susceptible person can be infected, and (b) an infected individual may become resistant if his immune system fights off the infection and confers resistance, or if the individual dies by the infection. These interactions can be modeled by a reaction network comprised of $N = 3$ species (S, I, and R) that interact through the following $M = 2$ reactions:

$$\begin{aligned} X_1 + X_2 &\to 2X_2 \\ X_2 &\to X_3 \ , \end{aligned} \tag{2.2.4}$$

where $X_1 = \mathrm{S}$, $X_2 = \mathrm{I}$ and $X_3 = \mathrm{R}$. In this case,

$$\mathbb{V} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbb{V}' = \begin{bmatrix} 0 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbb{S} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}. \tag{2.2.5}$$

We can now assume that the probability of a randomly selected susceptible individual at time t to become infected by a randomly selected infectious individual during an infinitesimally small time interval $[t, t + dt)$ is proportional to dt, with proportionality factor κ_1 that does not depend on the particular individuals involved. Moreover, we can assume that the probability of a randomly selected infected individual at time t to recover or die from the disease during $[t, t + dt)$ is also proportional to dt, with proportionality factor κ_2 that does not depend on the particular infected individual. Then, the previous interactions lead to a Markovian reaction network with mass-action propensity functions given by [17]

$$\pi_1(x_1, x_2, x_3) = \kappa_1 x_1 x_2 \quad \text{and} \quad \pi_2(x_1, x_2, x_3) = \kappa_2 x_2, \tag{2.2.6}$$

where x_1, x_2, x_3 are the populations of susceptible, infectious, and resistant individuals, respectively.

We can use the previous 3-species/2-reactions motif, given by Eq. (2.2.4), to construct more complex Markovian reaction networks that model the spread of an infectious disease in a population of individuals grouped into classes (e.g., households, work spaces, cities, etc.); see [18].

We may group, for example, individuals into two classes, those living in Baltimore and Philadelphia, and give each class its own distinct set of variables, namely X_1, X_2, X_3, for susceptible, infected, and resistant individuals in Baltimore, as well as X_4, X_5, X_6, for susceptible, infected, and resistant individuals in Philadelphia. Each class will be characterized by the previous 3-species/2-reactions motif, resulting in the following four reactions:

$$\begin{aligned} X_1 + X_2 &\to 2X_2 \\ X_2 &\to X_3 \\ X_4 + X_5 &\to 2X_5 \\ X_5 &\to X_6 \ . \end{aligned} \tag{2.2.7}$$

In this case however there is also a flow (by air, road, or rail) of individuals between the two different cities, which we can model by using the following six reactions:

$$\begin{aligned} X_1 &\to X_4 \\ X_4 &\to X_1 \\ X_2 &\to X_5 \\ X_5 &\to X_2 \\ X_3 &\to X_6 \\ X_6 &\to X_3 \ . \end{aligned} \tag{2.2.8}$$

The propensity functions associated with these new reactions will be proportional to the population of the input species, with the proportionality factor being the specific probability rate constant of an individual traveling from one city to the other. In this fashion, we can build complex Markovian reaction network models for epidemiological dynamics that are more realistic and more predictive than traditional deterministic models.

Likewise, new reactions may be incorporated into the epidemiological network to account for additional transitions between states. For instance, if we assume that a vaccine is available, then we must include the reaction $X_1 \to X_3$ in the formulation. Vital dynamics (i.e., births and

deaths) may also be included in this fashion. For example, if infants born at a fixed rate are always susceptible, then the reaction $\emptyset \to X_1$ must be included in the system. Finally, one may consider social networks on which epidemiological networks reside. Specifically, age stratification in the population [16], or the scale-free structure of social/sexual networks [7], may be handled in a manner similar – albeit not identical – to the aforementioned geographic considerations.

2.2.3 Neural networks

A discussion on reaction networks cannot be complete without mentioning biological neural networks. With 100 billion or more neurons in the human brain connected by 100-500 trillion synapses, there is no other reaction network that can compete in size and complexity.

There is a large body of literature surrounding the modeling and analysis of biological neural networks. As an example, we consider a Markovian reaction model for neural networks recently proposed in [19] that is intuitive enough for novices in neurobiology to comprehend and yet rich enough to be a viable candidate for understanding many features of this preeminent reaction network. The model consists of L neurons, with each neuron being in either a quiescent or an active state. Let X_{2l-1} and X_{2l} denote a quiescent or active neuron l, respectively. We can assign the following two reactions to the l-th neuron in the network:

$$\begin{aligned} X_{2l-1} + \sum_{l' \neq l} \nu_{l'l} X_{2l'} &\to X_{2l} + \sum_{l' \neq l} \nu_{l'l} X_{2l'} \\ X_{2l} &\to X_{2l-1}, \end{aligned} \tag{2.2.9}$$

where ν_{ij} measures the synaptic weight between neurons i and j, with a positive value indicating an excitatory synapsis and a negative value indicating an inhibitory synapsis. Note that the first reaction models transition of the l-th neuron from the quiescent to the active state, which is assumed to be influenced by appropriately weighted active neurons $X_{2l'}$, $l' \neq l$, in the network [see Eq. (2.2.10) below] that act as "catalysts." On the other hand, the second reaction models transition of the neuron from the active to the quiescent state, which is assumed to occur constitutively. As a consequence, we obtain a reaction network with $N = 2L$ species and $M = 2L$ reactions.

We can describe this system by a $2L \times 1$ state vector $\boldsymbol{x}$ with binary-valued 0/1 elements x_{2l-1}, x_{2l} indicating the state of the l-th neuron (with 0 being quiescent and 1 being active). Due to the fact that a neuron must be either quiescent or active, the state variables must satisfy the "mass conservation" relationships $x_{2l-1} + x_{2l} = 1$, for $l = 1, 2, \ldots, L$. It has been suggested in [19] that the probability of the l-th neuron becoming active during an infinitesimally small time interval $[t, t+dt)$, given that the neuron is quiescent at time t, can be taken to be $x_{2l-1}[\phi_l(\boldsymbol{x}) > 0] \tanh[\phi_l(\boldsymbol{x})]dt + o(dt)$, where $[a > 0]$ is the Iverson bracket and ϕ_l is the net synaptic input to the l-th neuron, given by

$$\phi_l(\boldsymbol{x}) = \sum_{l' \neq l} \nu_{l'l} x_{2l'} + \eta_l, \tag{2.2.10}$$

with η_l being an external input to the neuron. The term x_{2l-1} ensures that the neuron becomes active within $[t, t + dt)$ only when it is quiescent at time t. As a consequence, the propensity of the first reaction in Eq. (2.2.9) will be given by

$$\pi_{2l-1}(\boldsymbol{x}) = x_{2l-1}[\phi_l(\boldsymbol{x}) > 0] \tanh[\phi_l(\boldsymbol{x})], \tag{2.2.11}$$

and therefore depends on the synaptic inputs from neurons connected to the l-th neuron and any external input to that neuron. On the other hand, if we assume that the l-th neuron decays from an active to a quiescent state at a constant rate γ_l, then the propensity of the second reaction will be given by

$$\pi_{2l}(\boldsymbol{x}) = \gamma_l x_{2l}, \tag{2.2.12}$$

where the term x_{2l} ensures that the neuron becomes inactive within $[t, t + dt)$ only when it is active at time t.

2.3 Solving the master equation

Although the algebraic form of the master equations (2.1.6) and (2.1.8) is simple, solving these equations [i.e., calculating the probabilities $p_{\boldsymbol{Z}}(\boldsymbol{z}; t)$ and $p_{\boldsymbol{X}}(\boldsymbol{x}; t)$ at each time $t > 0$] is a difficult task in general. Many methods have been proposed in the literature to address this problem, which

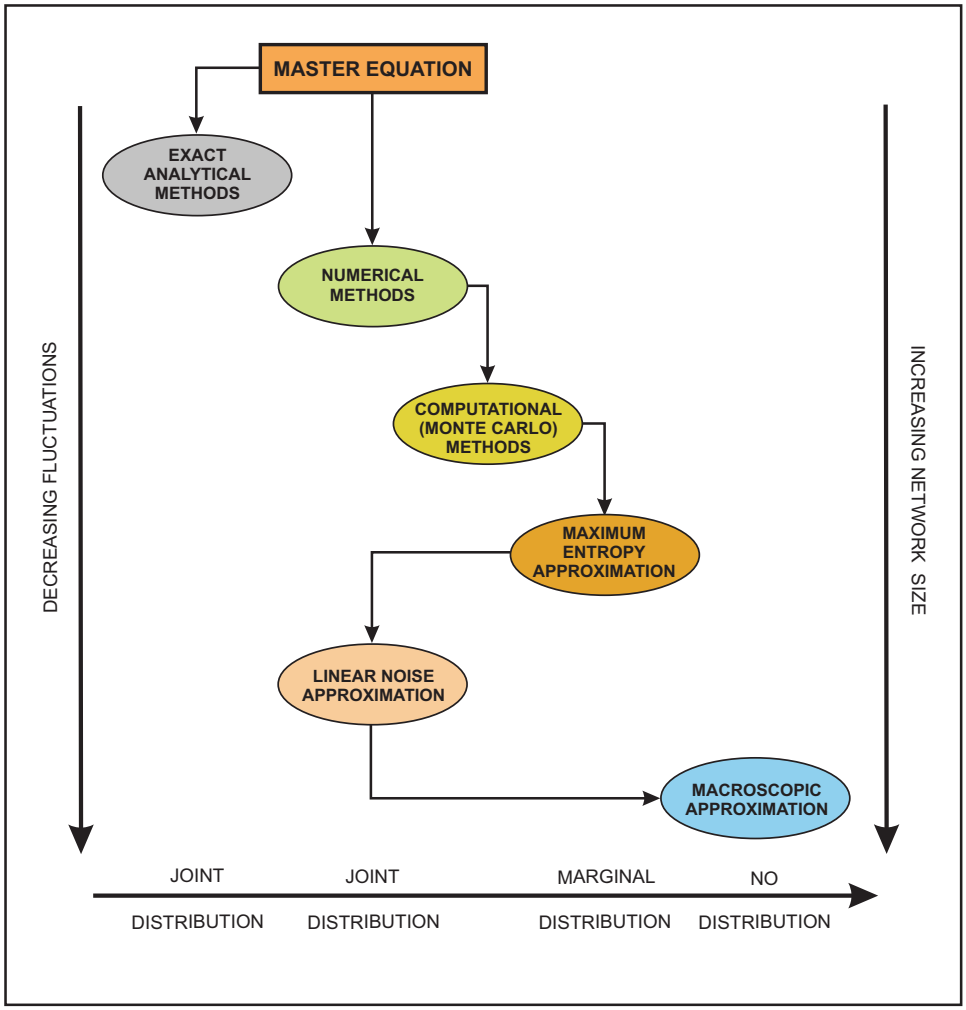

Figure 2.2: Six methods for solving the master equation. Some methods can be used to approximate the *joint* probability distributions of the DA and population processes while other methods can only be used to approximate *marginal* distributions. Exact analytical solutions can be obtained only in special cases. Numerical methods are currently limited to small reaction networks. Large networks require use of a maximum entropy approximation scheme (not discussed in this thesis) or adoption of the linear noise approximation method as opposed to a computational method based on Monte Carlo sampling. For large reaction networks, the macroscopic approximation may be the only feasible choice. This approximation however can in general be trusted only at low fluctuation levels.

can be grouped into the six general categories depicted in Fig. 2.2. In the following, we discuss the most prominent techniques available to date. Whether a given method can be applied to a particular problem depends on the size and complexity of the reaction network at hand.

2.3.1 Exact analytical methods

Deriving exact analytical solutions for $p_{\boldsymbol{Z}}(\boldsymbol{z};t)$ and $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ is possible only in simple cases (e.g., see [20–29]). For example, an analytical solution for the master equation (2.1.8) can be derived in the case of a *linear* reaction network (i.e., a network with linear propensity functions). It has been shown in [25] that, for *closed* linear reaction networks (i.e., linear reaction networks with fixed net population), the solution of the master equation (2.1.8) is a multinomial distribution, provided that the initial joint distribution is also multinomial. Moreover, for *open* linear reaction networks (i.e., linear reaction networks with varying net population), the solution of the master equation (2.1.8) is a product Poisson distribution, provided that the initial joint distribution is also product Poisson (see also [27]). These results are special cases of a more general result derived in [28] according to which the probability distribution $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ of the population process in a linear reaction network with initial state $\boldsymbol{x}_0$ can be expressed as the convolution of multinomial and product Poisson distributions with time-dependent parameters that evolve according to well-defined systems of first-order linear differential equations (see also [26]).

2.3.2 Numerical methods

Substantial effort has been focused recently on approximately solving the master equation (2.1.8) using numerical techniques. Although the methods developed so far show promise for addressing this problem, they are mostly limited to relatively small reaction networks. For this reason, we only provide a brief discussion here. The interested reader can find details in the references.

The master equation (2.1.8) can be expressed as a linear system of *coupled* first-order differential equations, given by

$$\frac{d\boldsymbol{p}(t)}{dt} = \mathbb{P}\boldsymbol{p}(t), \quad t > 0, \tag{2.3.1}$$

where $\boldsymbol{p}(t)$ is a $K \times 1$ vector that contains the nonzero probabilities $p_{\boldsymbol{X}}(\boldsymbol{x};t)$, $\boldsymbol{x} \in \mathcal{X}$, of the population process $\boldsymbol{X}(t)$ and $\mathbb{P}$ is a large $K \times K$ *sparse* matrix whose structure can be inferred directly from the

master equation. When the columns of the net stoichiometric matrix $\mathbb{S}$ are all different from each other, the only nonzero elements of the i-th column of $\mathbb{P}$ are M off-diagonal elements with values given by $\pi_m(\boldsymbol{x}_i)$, and the diagonal element, whose value is given by $-\sum_{m=1}^{M}\pi_m(\boldsymbol{x}_i)$, where $M \ll K$ is the number of reactions. If we assume that the cardinality K of the state-space $\mathcal{X}$ is finite, then we can calculate the probabilities $p_{\boldsymbol{x}}(\boldsymbol{x};t)$ by solving Eq. (2.3.1), in which case

$$\boldsymbol{p}(t) = \exp(t\mathbb{P})\,\boldsymbol{p}(0), \quad \text{for } t \geq 0. \tag{2.3.2}$$

This simple idea has led to a numerical technique, proposed in [30], for approximately solving the master equation known as *finite state projection* (FSP). This method requires an appropriate truncation of the state-space to determine the smallest possible set $\mathcal{X}$ and development of a computationally feasible algorithm for calculating the matrix exponential in Eq. (2.3.2).

Although a number of methods are available for computing matrix exponentials (e.g., see [31]), we briefly discuss here a popular technique known as *Krylov subspace approximation* (KSA) method [32, 33]. For a sufficiently small time step $\tau > 0$, this is the best available method for approximating the vector $\boldsymbol{p}(t+\tau) = \exp(\tau\mathbb{P})\,\boldsymbol{p}(t)$, when $\mathbb{P}$ is a large and sparse matrix. This is done by using a polynomial series expansion of the form:

$$\widehat{\boldsymbol{p}}(t+\tau) = c_0\boldsymbol{p}(t) + c_1\tau\mathbb{P}\boldsymbol{p}(t) + \cdots + c_{K_0-1}(\tau\mathbb{P})^{K_0-1}\boldsymbol{p}(t), \tag{2.3.3}$$

where the coefficients $c_0, c_1, \ldots, c_{K_0-1}$ are estimated by minimizing the least-squares error $||\boldsymbol{p}(t+\tau) - \widehat{\boldsymbol{p}}(t+\tau)||_2^2$. It turns out that the optimal K_0-th order polynomial approximation of $\boldsymbol{p}(t+\tau)$ is a point in the K_0-dimensional Krylov subspace $\mathcal{K}(t) = \text{span}\big\{\boldsymbol{p}(t), \tau\mathbb{P}\boldsymbol{p}(t), \ldots, (\tau\mathbb{P})^{K_0-1}\boldsymbol{p}(t)\big\}$. This element can be approximated by

$$\widehat{\boldsymbol{p}}(t+\tau) := ||\boldsymbol{p}(t)||_2\mathbb{V}(t)\exp\{\tau\mathbb{H}(t)\}\,\boldsymbol{e}_1, \tag{2.3.4}$$

where $\mathbb{V}(t)$ is a $K \times K_0$ matrix whose columns form an orthonormal basis for the Krylov subspace $\mathcal{K}(t)$ and $\mathbb{H}(t)$ is a $K_0 \times K_0$ Hessenberg matrix (upper triangular with an extra subdiagonal), both computed by the well-known Arnoldi procedure [33]. Finally, $\boldsymbol{e}_1$ is the first column of the $K_0 \times K_0$ identity matrix.

The KSA method reduces the problem of calculating the exponential of a large and sparse $K \times K$ matrix $\mathbb{P}$ to the problem of calculating the exponential of the much smaller and dense $K_0 \times K_0$ matrix $\mathbb{H}$ ($K_0 \ll K$, with $K_0 = 30$–50 being sufficient for many applications). Computation of the reduced size problem can be done by standard methods, such as a Chebyshev or Padé approximation [31–33]. Note that we can recursively estimate the solution $\boldsymbol{p}(t)$ in Eq. (2.3.2) at some time t_j by

$$\widehat{\boldsymbol{p}}(t_j) = \exp\{(t_j - t_{j-1})\mathbb{P}\}\,\widehat{\boldsymbol{p}}(t_{j-1}) = ||\widehat{\boldsymbol{p}}(t_{j-1})||_2 \mathbb{V}(t_{j-1}) \exp\{(t_j - t_{j-1})\mathbb{H}(t_{j-1})\}\boldsymbol{e}_1, \tag{2.3.5}$$

for $j = 1, 2, \ldots$, where $\widehat{\boldsymbol{p}}(0) = \boldsymbol{p}(0)$ and $0 = t_0 < t_1 < t_2 < \cdots$ is an increasing sequence of (not necessarily uniformly spaced) time points. These points are selected automatically, in conjunction with an appropriately designed error estimation procedure, to ensure stability and accuracy of the overall algorithm [32].

Unfortunately, and for most realistic reaction networks, $\mathcal{X}$ contains an extremely large number of states with non-negligible probability, thus making the practical implementation of FSP difficult. This is a direct consequence of the fact that $\mathcal{X}$ contains $R_1 \times R_2 \times \cdots \times R_N$ distinct elements, where R_n is an assumed maximum copy number of the n-th species. A number of approaches have been proposed in the literature to address this problem [34–42]. Although some approaches perform well, most are limited to small reaction networks. It turns out that the most difficult issue associated with these methods is solving the resulting system of differential equations, which is usually prohibitively large.

We should point out here that another numerical approach has been recently proposed in the literature that also attempts to address the previous problem [43, 44]. The method is based on representing the probability mass function of the population process by an appropriately chosen wavelet decomposition scheme whose basis elements and the associated wavelet coefficients are being adaptively updated in time by solving a much smaller system of linear equations. Although preliminary results indicate that the method works well, it is not clear at this point whether it can

be efficiently used to evaluate population probabilities in reaction networks containing more than a few reactions and species.

2.3.3 Computational methods

Numerical approaches for solving the master equation are not practical when the reaction network contains many reactions and species. In this case, computational techniques, based on Monte Carlo sampling, can be used to approximately evaluate the statistical behavior of the network. If, by simulation, we generate L sample trajectories $\{\boldsymbol{z}^{(l)}(t), t > 0\}$, $l = 1, 2, \ldots, L$, of the DA process $\{\boldsymbol{Z}(t), t > 0\}$, then we can estimate the dynamics of its moments, such as of the means $\{\mu_{\boldsymbol{z}}(m;t) := \mathrm{E}[Z_m(t)], t > 0\}$ and covariances $\{c_{\boldsymbol{z}}(m, m';t) := \mathrm{cov}[Z_m(t), Z_{m'}(t)], t > 0\}$, by using the following Monte Carlo estimators:

$$\widehat{\mu}_{\boldsymbol{z}}(m;t) \;=\; \frac{1}{L}\sum_{l=1}^{L} z_m^{(l)}(t), \tag{2.3.6}$$

$$\widehat{c}_{\boldsymbol{z}}(m, m';t) \;=\; \frac{1}{L-1}\sum_{l=1}^{L}\left[z_m^{(l)}(t) - \widehat{\mu}_{\boldsymbol{z}}(m;t)\right]\left[z_{m'}^{(l)}(t) - \widehat{\mu}_{\boldsymbol{z}}(m';t)\right]. \tag{2.3.7}$$

Moreover, we can estimate the probability distribution $p_{\boldsymbol{z}}(\boldsymbol{z};t)$ by using

$$\widehat{p}_{\boldsymbol{z}}(\boldsymbol{z};t) = \frac{1}{L}\sum_{l=1}^{L}\Delta(\boldsymbol{z}^{(l)}(t) - \boldsymbol{z}), \tag{2.3.8}$$

where $\Delta(\boldsymbol{z})$ is the Kronecker delta function. Due to the simple relationship between the DA and population processes given by Eq. (2.1.4), we can use similar estimators to approximate the dynamic evolution of the corresponding population statistics.

Unfortunately, to obtain sufficiently accurate Monte Carlo estimates, we need a large number of sample trajectories, which is computationally inefficient, especially when estimating high-order moments or probability distributions. As a matter of fact, when estimating probability distributions, the issue of efficiently sampling low probability events is crucial and becomes the main bottleneck for deriving accurate and computationally efficient Monte Carlo estimators. This problem can be addressed by developing computationally efficient approaches for sampling the master equation (2.1.6).

In the following, we discuss a number of methods available in the literature.

Exact sampling

The simplest way to draw samples from the master equation (2.1.6) is by using the *exact* algorithm of Gillespie [14,45–47]. By using simple probabilistic arguments, it has been shown in [14] that, given the system state $\boldsymbol{z}(t)$ at time t, the probability that the next reaction will occur at time $t+\tau+dt$ and that this will be the m-th reaction is given by $p_t(\tau,m)dt$, where (see Appendix A)

$$p_t(\tau,m)=\alpha_m(\boldsymbol{z}(t))\exp\left\{-\tau\sum_{m\in\mathcal{M}}\alpha_m(\boldsymbol{z}(t))\right\},\quad \text{for } \tau>0,\quad m\in\mathcal{M}. \tag{2.3.9}$$

As a consequence,

$$p_t(\tau,m)=\frac{\alpha_m(\boldsymbol{z}(t))}{\sum_{m'\in\mathcal{M}}\alpha_{m'}(\boldsymbol{z}(t))}\left[\sum_{m'\in\mathcal{M}}\alpha_{m'}(\boldsymbol{z}(t))\right]\exp\left\{-\tau\sum_{m\in\mathcal{M}}\alpha_m(\boldsymbol{z}(t))\right\}=r_t(m)e_t(\tau), \tag{2.3.10}$$

where

$$r_t(m):=\frac{\alpha_m(\boldsymbol{z}(t))}{\sum_{m'\in\mathcal{M}}\alpha_{m'}(\boldsymbol{z}(t))},\quad \text{for } m\in\mathcal{M}, \tag{2.3.11}$$

and

$$e_t(\tau):=\left\{\sum_{m\in\mathcal{M}}\alpha_m(\boldsymbol{z}(t))\right\}\exp\left\{-\tau\sum_{m\in\mathcal{M}}\alpha_m(\boldsymbol{z}(t))\right\},\quad \text{for } \tau>0, \tag{2.3.12}$$

which is an exponential distribution. This implies that the time of the next reaction and the index of the next reaction are statistically independent random variables with probability density and mass functions $e_t(\tau)$ and $r_t(m)$, respectively. We can therefore generate a trajectory $\{\boldsymbol{z}(t),t>0\}$ of the DA process by following two steps. First, given that the system is at state $\boldsymbol{z}(t)$ at time t, the time $t+\tau$ of the next reaction to occur can be determined by drawing a sample τ from the exponential distribution $e_t(\tau)$. We can then specify which reaction occurs at time $t+\tau$ by drawing a sample from the probability mass function $r_t(m)$ and by increasing the corresponding value of $\boldsymbol{z}$ by one.

Unfortunately, the Gillespie algorithm is computationally demanding, especially when applied to large and highly reactive systems, due to the fact that every single reaction event must be

faithfully simulated. As a consequence, calculating a typical realization of the DA process often requires a large number of samples to be drawn from the probability distributions given by Eq. (2.3.11) and Eq. (2.3.12), thus appreciably increasing computational complexity. Attempts in [48–50] to improve the computational efficiency of the Gillespie algorithm have produced sampling methods that significantly increase computational speed for large reaction networks. We refer the reader to [51–58] for alternative simulation algorithms designed to accelerate exact sampling of the master equation under certain conditions. Despite these efforts however, the previous methods are still inefficient, especially when used in conjunction with Monte Carlo estimation. For this reason, work has focused on developing approximate sampling techniques that appreciably reduce computational complexity by trading-off accuracy. We discuss some of these methods next.

Poisson leaping

The Markovian nature of the DA process $\boldsymbol{Z}(t)$ implies that [59, Theorem 5.8]:

$$Z_m(t) = P_m\left[\int_0^t \alpha_m(\boldsymbol{Z}(t'))\,dt'\right], \quad \text{for } t > 0, \quad m \in \mathcal{M}, \tag{2.3.13}$$

where P_m, $m \in \mathcal{M}$, are statistically independent Poisson random variables with unit rate. Moreover,

$$Z_m(t+\tau) = Z_m(t) + P_m\left[\int_t^{t+\tau} \alpha_m(\boldsymbol{Z}(t'))\,dt'\right], \quad \text{for } t > 0, \quad m \in \mathcal{M}, \tag{2.3.14}$$

for every $\tau > 0$, by virtue of the fact that a Poisson random variable with rate $\lambda_1 + \lambda_2$ can be written as the sum of two independent Poisson random variables with rates λ_1 and λ_2. As a consequence, we can use Eq. (2.3.14) to construct a technique for approximately sampling the master equation which, under certain circumstances, turns out to be accurate and computationally efficient. In particular, we will assume that a time step τ can be found so that, for every $j = 0, 1, \ldots$, the occurrence of reactions within the time interval $[j\tau, (j+1)\tau)$ does not appreciably affect the propensity functions α_m, $m \in \mathcal{M}$. In this case, Eq. (2.3.14) becomes

$$Z_m((j+1)\tau) \simeq Z_m(j\tau) + P_m\left[\alpha_m(\boldsymbol{Z}(j\tau))\tau\right], \quad \text{for } j = 0, 1, \ldots, \quad m \in \mathcal{M}, \tag{2.3.15}$$

initialized by $Z_m(0) = 0$, for every $m \in \mathcal{M}$.

We can now use Eq. (2.3.15) to approximately sample the master equation in an iterative fashion. Starting with zero DA values at time zero, we can approximate the DA process at time τ by setting $\widehat{z}_m(\tau) = p_m^{(0)}$, for every $m \in \mathcal{M}$, where $p_m^{(0)}$ is a sample drawn from the Poisson distribution with rate $\alpha_m(\mathbf{0})\tau$. Then, we can approximate the DA process at time 2τ by setting $\widehat{z}_m(2\tau) = \widehat{z}_m(\tau) + p_m^{(1)}$, for every $m \in \mathcal{M}$, where $p_m^{(1)}$ is a sample drawn from the Poisson distribution with rate $\alpha_m(\widehat{\boldsymbol{z}}(\tau))\tau$, and so on.

By using Eq. (2.3.15), we expect to obtain accurate samples of the DA process, provided that we can find a time step τ for which the required *leap* condition

$$\int_{j\tau}^{(j+1)\tau} \alpha_m(\boldsymbol{Z}(t'))\, dt' \simeq \alpha_m(\boldsymbol{Z}(j\tau))\tau \tag{2.3.16}$$

is satisfied. We would like this value to be as large as possible so that the resulting method is appreciably faster than exact sampling. Practical considerations however dictate that τ must not be very large, otherwise the method may inaccurately estimate the numbers of reactions occurring during the time intervals $[j\tau, (j+1)\tau)$, which may lead to negative species populations. This may not be appropriate in certain types of networks, such as biochemical reaction networks.

The problem of determining the largest value of τ so that the leap condition given by Eq. (2.3.16) is satisfied has been addressed in [47,60–62]. The procedure developed in [62] is accurate, easy to code, and results in faster implementation than the methods proposed in [60, 61]. To avoid negative populations, it has been suggested in [63–65] to approximate the Poisson distribution by a binomial distribution. The main rationale behind this choice is that the maximum number of occurrences produced by a binomial distribution is always bounded and easily controlled by one of the two parameters used to specify the distribution. This however is not true for the Poisson distribution, which can produce an unreasonably large number of occurrences within a small time interval (a Poisson random variable takes values between 0 and ∞) that can falsely result in negative populations. Some improvements of the original Poisson leaping methods can be found in [66–70].

It turns out that we can still use a Poisson distribution for the occurrence of reactions and always guarantee nonnegative populations. This has been recognized in [47,71], in which a sampling method has been proposed that is easier to implement than binomial leaping and is more accurate in general than the original Poisson leaping technique. An improved version of this approach, which employs a post-leap check to improve sampling accuracy, can be found in [72].

Gaussian leaping

In addition to the leap condition given by Eq. (2.3.16), the expected number $\alpha_m(\boldsymbol{Z}(j\tau))\tau$ of occurrences of the m-th reaction during the time interval $[j\tau,(j+1)\tau)$ is almost surely large compared to one [i.e., $\alpha_m(\boldsymbol{Z}(j\tau))\tau \gg 1$ with probability one]. We can then approximate the Poisson distribution $P_m\left[\alpha_m(\boldsymbol{Z}(j\tau))\tau\right]$ in Eq. (2.3.15) by a normal distribution with mean and variance given by $\alpha_m(\boldsymbol{Z}(j\tau))\tau$. In this case, the DA process $\boldsymbol{Z}(t)$ will satisfy the following equations [10,14,45,46]:

$$Z_m((j+1)\tau) \simeq Z_m(j\tau) + \alpha_m(\boldsymbol{Z}(j\tau))\tau + \sqrt{\alpha_m(\boldsymbol{Z}(j\tau))\tau}\, G_m^{(j)}, \quad \text{for } j=0,1,\ldots,\ m\in\mathcal{M}, \quad (2.3.17)$$

initialized by $Z_m(0)=0$, for every $m\in\mathcal{M}$, where $\{G_m^{(j)}, j=0,1,\ldots, m\in\mathcal{M}\}$ are mutually uncorrelated standard normal random variables. We can now use Eq. (2.3.17) to approximately sample the master equation in an iterative fashion. Starting with zero DA values at time zero, we can approximate the DA process at time τ by setting $\widehat{z}_m(\tau) = \alpha_m(\boldsymbol{0})\tau + \sqrt{\alpha_m(\boldsymbol{0})\tau}\, g_m^{(0)}$, for every $m\in\mathcal{M}$, where $g_m^{(0)}$, $m\in\mathcal{M}$, are samples independently drawn from the standard normal distribution. Then, we can approximate the DA process at time 2τ by setting $\widehat{z}_m(2\tau) = \widehat{z}_m(\tau) + \alpha_m(\widehat{\boldsymbol{z}}(\tau))\tau + \sqrt{\alpha_m(\widehat{\boldsymbol{z}}(\tau))\tau}\, g_m^{(1)}$, for every $m\in\mathcal{M}$, where $g_m^{(1)}$, $m\in\mathcal{M}$, are new samples independently drawn from the standard normal distribution, and so on.

The previous Gaussian leaping method results in faster sampling of the master equation since drawing samples from the standard normal distribution is usually more efficient than drawing samples from the Poisson distribution. Unfortunately, Gaussian leaping may result in crude approximations of the DA and population processes [13]. The main culprit is our difficulty in determining an appropriate time step τ so that the two required conditions mentioned above are *simultaneously*

satisfied. For example, we may try to reduce τ so that the propensity functions do not change appreciably during any time interval $[j\tau, (j+1)\tau)$, thus satisfying the leap condition. However, if the reaction network contains "slow" reactions (a situation that appears often in practice), these reactions will occur infrequently during $[j\tau, (j+1)\tau)$, and the second condition will be violated. Note that, in sharp contrast to Poisson leaping that always produces integer-valued DA trajectories, Gaussian leaping will produce DA trajectories that are real-valued. Moreover, and similarly to Poisson leaping, Gaussian leaping may produce reaction occurrences within $[j\tau, (j+1)\tau)$ that may result in negative species populations (see also the discussion in pp. 65-71 of [73]).

2.3.4 Linear noise approximation

In certain circumstances, the joint probability distributions of the DA and population processes can be well approximated by multivariate normal distributions. To see why this is true, we will assume the existence of a system parameter Ω that measures the relative size of stochastic fluctuations in a Markovian reaction network, such that fluctuations are small for large Ω. This is motivated by the fact that, in chemical reaction systems, stochastic fluctuations gradually diminish as the system approaches the *thermodynamic limit* at which the population of each species and the system volume approach infinity in a way that the concentrations remain fixed. In the following, we denote the thermodynamic limit by $\Omega \to \infty$ and make explicit the dependance of various quantities on Ω when necessary.

It is intuitive to expect that the probability of a reaction to occur within an infinitesimally small time interval $[t, t+dt)$ depends on the "density" $\boldsymbol{x}(t;\Omega)/\Omega$ of the population process at time t and that this probability does not change when Ω varies as long as the population densities remain fixed [1]. This implies that the propensity functions π_m must satisfy $\pi_m(\boldsymbol{x};\Omega) = \widetilde{\pi}_m(\boldsymbol{x}/\Omega)$, where $\widetilde{\pi}_m$ does not depend on Ω. To be more general, we may also add a term $\Omega^{-1}\widetilde{\pi}'_m(\boldsymbol{x}/\Omega)$, in which case we would like $\pi_m(\boldsymbol{x};\Omega) = \widetilde{\pi}_m(\boldsymbol{x}/\Omega) + \Omega^{-1}\widetilde{\pi}'_m(\boldsymbol{x}/\Omega)$. Moreover, we can assume that $\widetilde{\pi}_m(\,\cdot\,)$ and

$\widetilde{\pi}'_m(\,\cdot\,)$ are analytic. Finally, we may allow an arbitrary positive factor $f(\Omega)$, such that

$$\pi_m(\boldsymbol{x};\Omega) = f(\Omega)\left[\widetilde{\pi}_m(\boldsymbol{x}/\Omega) + \Omega^{-1}\widetilde{\pi}'_m(\boldsymbol{x}/\Omega)\right]. \tag{2.3.18}$$

This implies the following scaling law for the propensity functions of the DA process:

$$\alpha_m(\boldsymbol{z};\Omega) = f(\Omega)\big[\widetilde{\alpha}_m(\boldsymbol{z}/\Omega) + \Omega^{-1}\widetilde{\alpha}'_m(\boldsymbol{z}/\Omega)\big], \quad \text{for } m \in \mathcal{M}, \tag{2.3.19}$$

where $\widetilde{\alpha}_m(\boldsymbol{z}/\Omega) := \widetilde{\pi}_m(\boldsymbol{x}_0/\Omega + \mathbb{S}\boldsymbol{z}/\Omega)$ and $\widetilde{\alpha}'_m(\boldsymbol{z}/\Omega) := \widetilde{\pi}'_m(\boldsymbol{x}_0/\Omega + \mathbb{S}\boldsymbol{z}/\Omega)$.

To proceed, we can make the following *ansatz*:

$$\widetilde{Z}_m(t;\Omega) = \zeta_m(t) + \frac{1}{\sqrt{\Omega}}\,\Xi_m(t), \quad \text{for } t > 0, \quad m \in \mathcal{M}, \tag{2.3.20}$$

where $\widetilde{Z}_m(t;\Omega)$ is the "density" $Z_m(t;\Omega)/\Omega$ of the DA process, $\Xi_m(t)$ is a noise component that quantifies the fluctuations associated with the DA process, and $\zeta_m(t)$ is a deterministic process that satisfies:

$$\frac{d\zeta_m(t)}{dt} = \widetilde{\alpha}_m(\boldsymbol{\zeta}(t)), \quad t > 0, \quad m \in \mathcal{M}, \tag{2.3.21}$$

initialized with $\zeta_m(0) = 0$. For each Ω, Eq. (2.3.20) decomposes the random DA density $\widetilde{Z}_m(t;\Omega)$ into a *macroscopic* (deterministic) component $\zeta_m(t)$ and an additive noise component $\Xi_m(t)$ that do not depend on Ω. Clearly, this equation is based on the premise that the fluctuations diminish to zero as fast as $\Omega^{-1/2}$. Eq. (2.3.20) must be justified. This can be done by a central limit theorem for the behavior of the probability density function of the DA density process $\widetilde{\boldsymbol{Z}}(t;\Omega)$, as $\Omega \to \infty$, similar to that shown in [74, 75] for the case of biochemical reaction networks.

By using Eqs. (2.3.19)–(2.3.21) and the Ω-expansion method of van Kampen, it can be shown (see Appendix A for a proof) that, for a sufficiently large Ω, the dynamic evolution of the probability density function $p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)$ of the noise vector $\boldsymbol{\Xi}(t)$ is approximately governed by the following *linear* Fokker-Planck equation [1, 76, 77]:

$$\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)}{\partial t} = \frac{1}{2}\sum_{m\in\mathcal{M}} \widetilde{\alpha}_m(\boldsymbol{\zeta}(t))\frac{\partial^2 p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)}{\partial \xi_m^2} - \sum_{m\in\mathcal{M}}\sum_{m'\in\mathcal{M}} \frac{\partial \widetilde{\alpha}_m(\boldsymbol{\zeta}(t))}{\partial \zeta_{m'}}\frac{\partial[\xi_{m'} p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial \xi_m}, \quad t > 0, \tag{2.3.22}$$

initialized with $p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};0) = \delta(\boldsymbol{\xi})$, where $\delta(\cdot)$ is the Dirac delta function. In this case, $\boldsymbol{\Xi}(t)$ will approximately be a normal random vector with zero mean and correlation matrix $\mathbb{C}_{\boldsymbol{\Xi}}(t)$ that satisfies the following Lyapunov matrix differential equation:

$$\frac{d\mathbb{C}_{\boldsymbol{\Xi}}(t)}{dt} = \mathbb{A}(t) + \mathbb{G}(t)\mathbb{C}_{\boldsymbol{\Xi}}(t) + \mathbb{C}_{\boldsymbol{\Xi}}(t)\mathbb{G}^T(t), \quad t > 0, \tag{2.3.23}$$

initialized with $\mathbb{C}_{\boldsymbol{\Xi}}(0) = \mathbb{O}$, where $\mathbb{O}$ is the null matrix. In this equation, $\mathbb{A}(t)$ and $\mathbb{G}(t)$ are two $M \times M$ matrices with elements

$$a_{m,m'}(t) = \widetilde{\alpha}_m(\boldsymbol{\zeta}(t))\,\Delta(m-m') \quad \text{and} \quad g_{m,m'}(t) = \frac{\partial \widetilde{\alpha}_m(\boldsymbol{\zeta}(t))}{\partial \zeta_{m'}}\,, \tag{2.3.24}$$

respectively, where $\Delta(m)$ is the Kronecker delta function. As a consequence, and for sufficiently large Ω, we can approximate the probability distribution $p_{\widetilde{\boldsymbol{Z}}}(\widetilde{\boldsymbol{z}};t)$ of the DA density process by a multivariate normal probability density function with mean $\boldsymbol{\zeta}(t)$, predicted by the macroscopic equations (2.3.21), and covariance matrix $\mathbb{C}_{\boldsymbol{\Xi}}(t)/\Omega$, predicted by the Lyapunov equation (2.3.23). Due to Eq. (2.1.4), this also allows us to approximate the probability distribution $p_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}};t)$ of the population density process $\widetilde{\boldsymbol{X}}(t;\Omega) := \boldsymbol{X}(t;\Omega)/\Omega$ by a multivariate normal probability density function with mean $\boldsymbol{x}_0/\Omega + \mathbb{S}\boldsymbol{\zeta}(t)$ and covariance matrix $\mathbb{S}\mathbb{C}_{\boldsymbol{\Xi}}(t)\mathbb{S}^T$. Since $\boldsymbol{Z}(t;\Omega) = \Omega\widetilde{\boldsymbol{Z}}(t;\Omega)$, we can also approximate the probability distribution $p_{\boldsymbol{Z}}(\boldsymbol{z};t)$ of the DA process with a multivariate normal distribution, with mean $\Omega\boldsymbol{\zeta}(t)$ and covariance matrix $\Omega\mathbb{C}_{\boldsymbol{\Xi}}(t)$, whereas, we can approximate the probability distribution $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ of the population process with a multivariate normal distribution with mean $\boldsymbol{x}_0 + \Omega\mathbb{S}\boldsymbol{\zeta}(t)$ and covariance matrix $\Omega\mathbb{S}\mathbb{C}_{\boldsymbol{\Xi}}(t)\mathbb{S}^T$.

Because fluctuations in the reaction network are governed by the linear "signal-plus-noise" model given by Eq. (2.3.20), the previous method is known as *linear noise approximation* (LNA). Its use requires specification of an appropriate fluctuation size parameter Ω, such that Eq. (2.3.20) is satisfied, and a sufficiently large value for this parameter so that the method produces a reasonable approximation of the true probability distributions $p_{\widetilde{\boldsymbol{Z}}}(\widetilde{\boldsymbol{z}};t)$ and $p_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}};t)$. Implementation of the method requires that we separately solve the system of M first-order differential equations (2.3.21) and the system of $M(M+1)/2$ first-order differential equations (2.3.23). The LNA method decouples

the computation of the means from the computation of the covariances. It turns out that the LNA method is substantially faster than Monte Carlo estimation and can be used to provide a rapid assessment of the statistical behavior of some Markovian reaction networks [13]. This method has already been used to study biochemical reaction networks [78–85], epidemiological networks [17], ecological networks [86, 87], social networks [88], and neural networks [19, 89].

2.3.5 Macroscopic approximation

For large nonlinear reaction networks, the LNA method can become computationally intractable, since evaluation of the covariances requires solving a system of $\mathcal{O}(M^2)$ differential equations. If that turns out to be the case, then the only option left to characterize the dynamic behavior of the reaction network is in terms of DA or population densities by using, for example, the macroscopic (fluctuation-free) system of M differential equations given by Eq. (2.3.21). As a matter of fact, Eq. (2.3.20) implies that, for any $t > 0$, the DA density process $\widetilde{Z}_m(t;\Omega)$ converges in distribution to $\zeta_m(t)$ as $\Omega \to \infty$.

Similarly to the DA density process, the population density process $\widetilde{\boldsymbol{X}}(t;\Omega)$ converges in distribution, as $\Omega \to \infty$, to the deterministic process $\boldsymbol{\chi}(t)$ that satisfies the following macroscopic equations:

$$\frac{d\chi_n(t)}{dt} = \sum_{m\in\mathcal{M}} s_{nm}\widetilde{\pi}_m(\boldsymbol{\chi}(t)), \quad t > 0, \quad n \in \mathcal{N}, \tag{2.3.25}$$

where $\widetilde{\pi}_m(\widetilde{\boldsymbol{x}}) := \Omega^{-1}\pi_m(\Omega\widetilde{\boldsymbol{x}})$, provided that these equations are initialized with the same condition as the master equation (2.1.8). This is clearly true at finite times. It is also true in the limit as $t \to \infty$, provided that the macroscopic equations (2.3.25) have a *unique* asymptotically stable stationary solution that is independent of the initial state [1, 90].

2.4 Mesoscopic (probabilistic) behavior

When studying Markovian reaction networks, an important goal is to derive mathematical properties of the dynamic behavior of the probability distribution of the system state and investigate the existence, uniqueness, and stability of a stationary solution of the underlying master equation. This can be done by using a *mesoscopic* description of the network in terms of the population probabilities $\{p_{\boldsymbol{x}}(\boldsymbol{x};t), \boldsymbol{x} \in \mathcal{X}\}$, for $t \geq 0$. To avoid mathematical subtleties, which are outside the scope of this section, we assume that the cardinality of the population state-space $\mathcal{X}$ is *finite*. Most results however can be extended to the case of countable state-spaces.

To derive a stationary solution of the master equation (2.1.8), we must solve the system of K linear equations $\mathbb{P}\boldsymbol{p} = 0$; recall Eq. (2.3.1). Since the elements of each column of matrix $\mathbb{P}$ add to zero, its rows are linearly dependent and, therefore, the rank of $\mathbb{P}$ will be less than K. As a consequence, the system of equations $\mathbb{P}\boldsymbol{p} = 0$ will have at least one nontrivial solution. Unfortunately, this result does not tell us how many nontrivial solutions exist and which ones are valid probability distributions; i.e., which solutions satisfy the necessary constraints

$$0 \leq p_k \leq 1, \quad \text{for } k = 1, 2, \ldots, K, \quad \text{and} \quad \sum_{k=1}^{K} p_k = 1. \tag{2.4.1}$$

In the following, we first focus our interest on *irreducible* Markovian reaction networks. This type of networks are defined by the property that, for any pair $(\boldsymbol{x}, \boldsymbol{x}')$ of population states, there exists at least one sequence of reactions that takes the system from state $\boldsymbol{x}$ to state $\boldsymbol{x}'$ – these states are said to be *communicating*. By using a simple graph-theoretic analysis and Kirchhoff's theorem, it has been shown in [91] that an irreducible Markovian reaction network converges to a *unique* probability distribution $\bar{\boldsymbol{p}}$ at steady-state, which does not depend on the initial probability distribution $\boldsymbol{p}(0)$, such that $\mathbf{0} < \bar{\boldsymbol{p}} < \mathbf{1}$, where $\mathbf{0}$ and $\mathbf{1}$ are vectors whose elements are respectively all zero or one (see also [1]). As a consequence, in an irreducible Markovian reaction network, the population process can take any value in $\mathcal{X}$ at steady-state with *nonzero* probability.

On the other hand, the theory of systems of ordinary differential equations with constant

coefficients implies that, for a given initial probability distribution $\boldsymbol{p}(0)$, Eq. (2.3.1) is satisfied by a *unique* probability distribution $\boldsymbol{p}(t)$, which is analytic for all $0 \leq t < \infty$. Since the elements of each column of matrix $\mathbb{P}$ add to zero,

$$\frac{d[\mathbf{1}^T \boldsymbol{p}(t)]}{dt} = \mathbf{1}^T \frac{d\boldsymbol{p}(t)}{dt} = \mathbf{1}^T \mathbb{P}\, \boldsymbol{p}(t) = 0. \tag{2.4.2}$$

This result, together with the fact that $\mathbf{1}^T \boldsymbol{p}(0) = 1$, implies $\mathbf{1}^T \boldsymbol{p}(t) = 1$, for all $t \geq 0$. Unfortunately, it is not clear whether $\mathbf{0} \leq \boldsymbol{p}(t) \leq \mathbf{1}$, for every $t > 0$. It turns out however that, for an irreducible Markovian reaction network, $\mathbf{0} < \boldsymbol{p}(t) < \mathbf{1}$, for every $t > 0$ [91].

Eigenanalysis of matrix $\mathbb{P}$ can produce an analytical formula for the dynamic behavior of the unique probability distribution $\boldsymbol{p}(t)$. If λ_k, $k = 1, 2, \ldots, K$, are the eigenvalues of matrix $\mathbb{P}$, with corresponding right and left eigenvectors $\boldsymbol{r}_k$, $\boldsymbol{l}_k$, $k = 1, 2, \ldots, K$, respectively, then the solution to Eq. (2.3.1) is given by [31]

$$\boldsymbol{p}(t) = \exp(\mathbb{P}t)\, \boldsymbol{p}(0) = \sum_{k=1}^{K} c_k\, \boldsymbol{r}_k\, e^{\lambda_k t}, \quad \text{for } 0 \leq t \leq \infty, \tag{2.4.3}$$

where we assume here that the eigenvalues of $\mathbb{P}$ have the same algebraic and geometric multiplicity, an assumption satisfied by many Markovian reaction networks. In this case, the right and left eigenvectors are biorthogonal (i.e., $\boldsymbol{l}_k^T \boldsymbol{r}_{k'} = 0$, for every $k \neq k'$), which implies that the constants c_k are given by $c_k = \boldsymbol{l}_k^T \boldsymbol{p}(0) / \boldsymbol{l}_k^T \boldsymbol{r}_k$. As a consequence, we can use the eigenvalues and eigenvectors of $\mathbb{P}$ to analytically specify the entire mesoscopic behavior of a Markovian reaction network. Note that Eq. (2.4.3) and the fact that a non-trivial stationary solution always exists imply that at least one eigenvalue of $\mathbb{P}$ must be zero. For an irreducible Markovian reaction network, matrix $\mathbb{P}$ has only one zero eigenvalue, with the remaining $K - 1$ eigenvalues having negative real parts [91]. If we therefore assume that $\lambda_1 = 0$, then Eq. (2.4.3) implies that the stationary distribution will be given by $\bar{\boldsymbol{p}} = \boldsymbol{r}_1 / \|\boldsymbol{r}_1\|$, where $\boldsymbol{r}_1$ is the eigenvector corresponding to the zero eigenvalue and $\|\boldsymbol{r}\|$ is the ℓ_1-norm of vector $\boldsymbol{r}$. See [92, 93] for application of Eq. (2.4.3) to problems in epidemiology and computational biochemistry. Note however that computing the eigenvalues and eigenvectors of $\mathbb{P}$ is an extremely difficult task in general due to the large size of the underlying state-space.

Finally, the solution $\boldsymbol{p}(t)$, $t \geq 0$, of Eq. (2.3.1) turns out to be asymptotically stable with respect to $\overline{\boldsymbol{p}}$, in the sense that

$$\lim_{t\to\infty} D[\boldsymbol{p}(t), \overline{\boldsymbol{p}}] = 0, \tag{2.4.4}$$

where

$$D[\boldsymbol{p}, \boldsymbol{q}] := \sum_{k=1}^{K} p_k \ln \frac{p_k}{q_k} \geq 0 \tag{2.4.5}$$

is the Kullback-Leibler distance between the two probability distributions $\boldsymbol{p} = \{p_k, k = 1, 2, \ldots, K\}$ and $\boldsymbol{q} = \{q_k, k = 1, 2, \ldots, K\}$. As a matter of fact, $dD[\boldsymbol{p}(t), \overline{\boldsymbol{p}}]/dt \leq 0$, where equality is achieved only at steady-state.

To summarize, for a given initial probability vector $\boldsymbol{p}(0)$, the master equation associated with an *irreducible* Markovian reaction network has a *unique* and strictly positive solution $\mathbf{0} < \boldsymbol{p}(t) < \mathbf{1}$, $0 < t \leq \infty$. This solution is analytic for all $0 \leq t < \infty$, converges to a stationary distribution $\mathbf{0} < \overline{\boldsymbol{p}} < \mathbf{1}$ that does not depend on the initial probability distribution $\boldsymbol{p}(0)$, and is asymptotically stable with respect to $\overline{\boldsymbol{p}}$.

It is not in general easy to check whether a Markovian reaction network is irreducible. However, we often assume that a given Markovian reaction network is comprised of only reversible reactions (reactions which can occur in both directions with nonzero probability). This is a plausible assumption since, in principle, a transition between two physical states can occur in the reverse direction as well. In this case, and after appropriately ordering the states, we can cast matrix $\mathbb{P}$ into a block diagonal form with diagonal elements $\mathbb{P}^{(1)}, \mathbb{P}^{(2)}, \ldots, \mathbb{P}^{(J)}$, for some J, where each submatrix $\mathbb{P}^{(j)}$ is irreducible (when $J = 1$, matrix $\mathbb{P}$ is itself irreducible). The resulting Markovian reaction network is said to be *completely reducible* [1]. In this case, the original Markovian reaction network can be decomposed into J non-interacting subnetworks with non-overlapping state-spaces, which can be treated independently of each other. Each reaction subnetwork is characterized by *unique* dynamic and stationary solutions $\boldsymbol{p}^{(j)}(t)$, $\overline{\boldsymbol{p}}^{(j)}$, $j = 1, 2, \ldots, J$, which satisfy the aforementioned properties. However, the dynamic and stationary solutions of the original master equation are determined by the initial condition at time $t = 0$. If the master equation is initialized with a population vector in

the state-space of the j-th subnetwork, then its dynamic and stationary solution will be given by

$$\begin{bmatrix} \mathbf{0} \\ \vdots \\ \boldsymbol{p}^{(j)}(t) \\ \vdots \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{0} \\ \vdots \\ \bar{\boldsymbol{p}}^{(j)} \\ \vdots \\ \mathbf{0} \end{bmatrix},$$

respectively, where $\boldsymbol{p}^{(j)}(t)$ depends on the initial condition and $\bar{\boldsymbol{p}}^{(j)}$ does not.

A question that arises at this point is what happens when the Markovian reaction network contains irreversible reactions and matrix $\mathbb{P}$ is not irreducible. To get an idea, let us assume that, after appropriately ordering the states,

$$\mathbb{P} = \begin{bmatrix} \mathbb{P}^{(1)} & \mathbb{T}^{(1)} \\ \mathbb{O} & \mathbb{T} \end{bmatrix}, \tag{2.4.6}$$

where $\mathbb{O}$ denotes a null matrix, $\mathbb{P}^{(1)}$ and $\mathbb{T}$ are square matrices, $\mathbb{P}^{(1)}$ is irreducible, and at least one element of each column of $\mathbb{T}^{(1)}$ is strictly positive. The associated Markovian reaction network is said to be *incompletely reducible* [1]. Note that the nonzero elements of $\mathbb{T}^{(1)}$ correspond to nonreversible reactions. This is due to the fact that, if the propensity function of a forward reaction shows up in the (i, j) entry of matrix $\mathbb{P}$ which is in $\mathbb{T}^{(1)}$, then the propensity function of the reverse reaction will show up in the (j, i) entry of $\mathbb{P}$, which is zero. As a consequence, the reaction will necessarily be irreversible.

If we denote by $\boldsymbol{p}^{(1)}(t)$ and $\boldsymbol{p}^{(2)}(t)$ the probability distributions of the state vectors at time t, determined by the partition of the state-space suggested by the previous matrix $\mathbb{P}$, then the master equation results in the following two differential equations:

$$\frac{d\boldsymbol{p}^{(1)}(t)}{dt} = \mathbb{P}^{(1)}\boldsymbol{p}^{(1)}(t) + \mathbb{T}^{(1)}\boldsymbol{p}^{(2)}(t) \tag{2.4.7}$$

$$\frac{d\boldsymbol{p}^{(2)}(t)}{dt} = \mathbb{T}\boldsymbol{p}^{(2)}(t). \tag{2.4.8}$$

Clearly, one can solve the second equation independently from the first to obtain

$$\boldsymbol{p}^{(2)}(t) = \exp\left\{\mathbb{T}t\right\}\boldsymbol{p}^{(2)}(0). \tag{2.4.9}$$

On the other hand, the dynamic behavior of $\boldsymbol{p}^{(1)}$ is now driven by $\boldsymbol{p}^{(2)}(t)$ [unless $\boldsymbol{p}^{(2)}(0) = \mathbf{0}$], in which case $\boldsymbol{p}^{(1)}(t) = \exp\left\{\mathbb{P}^{(1)}t\right\}\boldsymbol{p}^{(1)}(0)$. Note however that

$$\frac{d[\mathbf{1}^T\boldsymbol{p}^{(2)}(t)]}{dt} = \mathbf{1}^T\frac{d\boldsymbol{p}^{(2)}(t)}{dt} = \mathbf{1}^T\mathbb{T}\boldsymbol{p}^{(2)}(t) = -\mathbf{1}^T\mathbb{T}^{(1)}\boldsymbol{p}^{(2)}(t) < 0, \tag{2.4.10}$$

provided that $\boldsymbol{p}^{(2)}(t) \neq \mathbf{0}$, since the elements of each column of matrix $\mathbb{P}$ add to zero and we have assumed that each column of matrix $\mathbb{T}^{(1)}$ contains at least one element that is strictly positive. Therefore, $\boldsymbol{p}^{(2)}(t)$ asymptotically becomes zero as $t \to \infty$. As a matter of fact, $\boldsymbol{p}^{(2)}(t)$ assigns probability mass over the *transient* states of the Markovian reaction network, as opposed to $\boldsymbol{p}^{(1)}(t)$ that assigns probability mass over the *persistent* states. In this case, and when matrix $\mathbb{P}^{(1)}$ is irreducible, the stationary solution of the master equation governing an incompletely reducible Markovian reaction network will be unique and given by the probability vector

$$\overline{\boldsymbol{p}} = \begin{bmatrix} \overline{\boldsymbol{p}}^{(1)} \\ \\ \mathbf{0} \end{bmatrix}, \tag{2.4.11}$$

where $\overline{\boldsymbol{p}}^{(1)}$ is the (unique) solution of the linear system of equations $\mathbb{P}^{(1)}\boldsymbol{p} = \mathbf{0}$.

In general, the population states in a Markovian reaction network can be classified into two distinct groups: *transient* and *persistent*. These states can be uniquely partitioned into non-overlapping sets T and P_j, $j = 1, 2, \ldots, J$, where T contains all transient states and P_j, $j = 1, 2, \ldots, J$, are irreducible sets containing persistent states with the additional property that, for every $j \neq j'$, each state in P_j does not communicate with any state in $P_{j'}$. By appropriately ordering the states,

we can write matrix $\mathbb{P}$ in the form

$$\mathbb{P} = \begin{bmatrix} \mathbb{P}^{(1)} & \mathbb{O} & \cdots & \mathbb{O} & \mathbb{T}^{(1)} \\ \mathbb{O} & \mathbb{P}^{(2)} & \cdots & \mathbb{O} & \mathbb{T}^{(2)} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbb{O} & \mathbb{O} & \cdots & \mathbb{P}^{(J)} & \mathbb{T}^{(J)} \\ \mathbb{O} & \mathbb{O} & \cdots & \mathbb{O} & \mathbb{T} \end{bmatrix}, \tag{2.4.12}$$

where $\mathbb{P}^{(j)}$ is a square irreducible matrix that characterizes how probability mass is dynamically distributed among the persistent states in P_j, $\mathbb{T}^{(j)}$ is a matrix that tells us how probability mass is transferred from the transient states in T to persistent states in P_j, $\mathbb{T}$ is a square matrix that characterizes how probability mass is dynamically distributed among the transient states in T, and $\mathbb{O}$ are null matrices. In this case, if the Markovian reaction network is initialized by a persistent state in P_j, then the stationary solution will be given by the probability vector

$$\boldsymbol{g}_j = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \overline{\boldsymbol{p}}^{(j)} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \tag{2.4.13}$$

where $\overline{\boldsymbol{p}}^{(j)}$ is the unique stationary distribution of the j-th irreducible Markovian reaction subnetwork characterized by matrix $\mathbb{P}^{(j)}$. However, if the network is initialized with the i-th transient state in T, then the stationary distribution $\overline{\boldsymbol{p}}_i$ (which now depends on i) will be given by a convex combination of the stationary distributions $\boldsymbol{g}_j$ above, with mixing coefficients μ_{ij}; i.e., we have that

$$\overline{\boldsymbol{p}}_i \ = \sum_{j=1}^{J} \mu_{ij} \boldsymbol{g}_j, \tag{2.4.14}$$

where

$$\mu_{ij} \geq 0 \qquad \text{and} \qquad \sum_{j=1}^{J} \mu_{ij} = 1. \tag{2.4.15}$$

As a matter of fact, Eq. (2.4.14) simply expresses the fact that the probability of a Markovian reaction network initialized with the i-th transient state in T to reach a persistent population state $\boldsymbol{x}$ in P_j at steady-state equals the probability μ_{ij} that the system will reach a persistent state in P_j at steady-state multiplied by the probability that this state will be $\boldsymbol{x}$. It can be shown (see Appendix A) that

$$\mu_{ij} = -\sum_{i' \in T} \sum_{j' \in P_j} [\mathbb{T}^{(j)}]_{j'i'} [\mathbb{T}^{-1}]_{i'i}, \tag{2.4.16}$$

where $[\mathbb{T}^{(j)}]_{j'i'}$ is the (j', i') element of matrix $\mathbb{T}^{(j)}$ and $[\mathbb{T}^{-1}]_{i'i}$ is the (i', i) element of the inverse of matrix $\mathbb{T}$.

To summarize, a fundamental property of the master equation (2.1.8) associated with a Markovian reaction network is that, when this equation is initialized with a persistent state, its solution converges to a unique stationary distribution that assigns positive probability only to the persistent states that communicate with the initial state. On the other hand, if the Markovian reaction network is initialized with a transient state, then its stationary distribution will be a convex combination of the distinct stationary distributions obtained by initializing the system with persistent states chosen from each individual irreducible set.

2.5 Potential energy landscape

To better understand what might happen at steady-state, let us assume that the master equation (2.1.8) has a unique stationary solution $\overline{p}_{\mathbf{X}}(\boldsymbol{x}) := \lim_{t \to \infty} p_{\boldsymbol{X}}(\boldsymbol{x}; t)$ that is independent of the initial state. In this case, the probability distribution $p_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}; t)$ of the population density process $\widetilde{\boldsymbol{X}}(t; \Omega) = \boldsymbol{X}(t; \Omega)/\Omega$ will be given by $p_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}; t) = \Omega p_{\boldsymbol{X}}(\Omega \widetilde{\boldsymbol{x}}; t)$ and will depend on the size parameter Ω in general. Let us define the function

$$V(\widetilde{\boldsymbol{x}}; \Omega) := -\frac{1}{\Omega} \ln \frac{\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})}{\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}_*)}, \tag{2.5.1}$$

where $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) := \lim_{t\to\infty} p_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}};t) = \Omega\overline{p}_{\boldsymbol{X}}(\Omega\widetilde{\boldsymbol{x}})$ is the steady-state distribution of the population density process and $\widetilde{\boldsymbol{x}}_*$ is a state at which the stationary probability distribution $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ attains its maximum value. Both $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ and $\widetilde{\boldsymbol{x}}_*$ depend on Ω but, for notational simplicity, we do not show this dependence. Note that $V(\widetilde{\boldsymbol{x}};\Omega) \geq 0$. Moreover, Eq. (2.5.1) implies that

$$\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) = \frac{1}{\zeta(\Omega)} \exp\Big\{-\Omega V(\widetilde{\boldsymbol{x}};\Omega)\Big\}, \tag{2.5.2}$$

where

$$\zeta(\Omega) := \sum_{\boldsymbol{u}} \exp\Big\{-\Omega V(\boldsymbol{u};\Omega)\Big\}. \tag{2.5.3}$$

In this case, $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ is a Gibbs distribution with "potential energy" function $V(\widetilde{\boldsymbol{x}};\Omega)$, "temperature" $1/\Omega$, and partition function $\zeta(\Omega)$. Clearly, $V(\widetilde{\boldsymbol{x}};\Omega)$ assigns minimum (zero) potential to the states of maximum probability at steady-state and infinite potential to the states of zero probability.

We will now assume that, close to the thermodynamic limit, the potential energy function $V(\widetilde{\boldsymbol{x}};\Omega)$ is an analytic function of Ω^{-1}. Then, a Taylor series expansion with respect to Ω^{-1} approximately results in

$$V(\widetilde{\boldsymbol{x}};\Omega) = V(\widetilde{\boldsymbol{x}};\infty) + \frac{1}{\Omega}\frac{\partial V(\widetilde{\boldsymbol{x}};\infty)}{\partial\Omega^{-1}} = V_0(\widetilde{\boldsymbol{x}}) + \frac{1}{\Omega}V_1(\widetilde{\boldsymbol{x}}), \tag{2.5.4}$$

for sufficiently large Ω, where

$$V_0(\widetilde{\boldsymbol{x}}) := V(\widetilde{\boldsymbol{x}};\infty) = -\lim_{\Omega\to\infty}\frac{1}{\Omega}\ln\frac{\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})}{\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}_*)} \geq 0, \tag{2.5.5}$$

and

$$V_1(\widetilde{\boldsymbol{x}}) := \frac{\partial V(\widetilde{\boldsymbol{x}};\infty)}{\partial\Omega^{-1}}. \tag{2.5.6}$$

As a consequence, Eqs. (2.5.2)–(2.5.4) approximately imply that

$$\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) = \frac{1}{\zeta(\Omega)} \exp\Big\{-\Omega V_0(\widetilde{\boldsymbol{x}}) - V_1(\widetilde{\boldsymbol{x}})\Big\}, \tag{2.5.7}$$

where the partition function is now given by

$$\zeta(\Omega) = \sum_{\boldsymbol{u}} \exp\Big\{-\Omega V_0(\boldsymbol{u}) - V_1(\boldsymbol{u})\Big\}. \tag{2.5.8}$$

If $\boldsymbol{\chi}(t)$ satisfies the macroscopic equations (2.3.25), then we can show (see Appendix A and [94]) that

$$\frac{dV_0(\boldsymbol{\chi}(t))}{dt} = \sum_{n\in\mathcal{N}} \frac{\partial V_0(\boldsymbol{\chi}(t))}{\partial \chi_n(t)} \frac{d\chi_n(t)}{dt} \leq 0, \tag{2.5.9}$$

provided that $V_0(\boldsymbol{\chi}(t)) < \infty$. As a consequence, the solution $\boldsymbol{\chi}(t)$ of the macroscopic equations (2.3.25) produces motion that never increases the value of the potential energy function V_0. If $\boldsymbol{\chi}'$ is a (strict) local minimum of V_0, we have that $V_0(\widetilde{\boldsymbol{x}}) > V_0(\boldsymbol{\chi}') \geq 0$, for every $\widetilde{\boldsymbol{x}} \in \mathcal{W}(\boldsymbol{\chi}')$, where $\mathcal{W}(\boldsymbol{\chi}')$ is a local neighborhood of $\boldsymbol{\chi}'$ that does not contain $\boldsymbol{\chi}'$. Then, Eq. (2.5.9) implies that V_0 is a (local) *Lyapunov function* for the macroscopic system and $\boldsymbol{\chi}'$ will be a (locally) *stable* solution of the macroscopic equations (2.3.25) in the sense of Lyapunov (i.e., the solution will always remain near $\boldsymbol{\chi}'$, provided that the macroscopic system is initialized by a state that is also near $\boldsymbol{\chi}'$) [95]. Moreover, if Eq. (2.5.9) is satisfied with strict inequality, unless $\boldsymbol{\chi}(t) = \boldsymbol{\chi}'$, then $\boldsymbol{\chi}'$ will be a (locally) *asymptotically stable* solution of the macroscopic equations (i.e., the solution will converge to $\boldsymbol{\chi}'$, provided that the macroscopic system is initialized by a state that is near $\boldsymbol{\chi}'$) [95]. Hence, a local minimum of V_0 must be a stable point of the macroscopic equations (2.3.25). It turns out that the inverse is also true. If $\boldsymbol{\chi}'$ is a (Lyapunov or asymptotically) stable equilibrium point of the macroscopic equations (2.3.25) that is not a local minimum of V_0, then the macroscopic equations, initialized by $\widetilde{\boldsymbol{x}}$ within a sufficiently small neighborhood of $\boldsymbol{\chi}'$ such that $V_0(\widetilde{\boldsymbol{x}}) < V_0(\boldsymbol{\chi}')$, will violate Eq. (2.5.9), since the system will need to increase the value of V_0 to get to $\boldsymbol{\chi}'$ from $\widetilde{\boldsymbol{x}}$. Therefore, there is a one-to-one correspondence between the local minima of V_0 and the stable points of the macroscopic equations (2.3.25). Similar results hold for the more general case when V_0 has a *regional* minima (i.e., compact sets of states with equal potential energy so that the energy increases as we move away from these states).

As a consequence of the previous arguments, we can view the multidimensional surface $V_0(\widetilde{\boldsymbol{x}})$ as a *potential energy landscape* [96–99] with the stable stationary states of the macroscopic equations (2.3.25) corresponding to *potential wells* (basins of attraction) associated with the minima

of V_0, separated by barriers corresponding to hills (unstable states) and saddles (transitional states – states on the potential energy surface from which stable states are equally accessible). Which path the macroscopic system takes along the potential energy landscape will depend on the initial condition. Initial conditions within a basin of attraction guarantee that the macroscopic dynamics will stay within the basin permanently. If the macroscopic system reaches a minimum of the potential energy landscape, then this minimum must be a stationary state of the macroscopic system since uphill motions are not possible. Thus, if the macroscopic system, characterized by Eq. (2.3.25), ever reaches a minimum of the potential energy landscape, it stays there forever.

We can now show (see Appendix A) that Eq. (2.5.7) and Eq. (2.5.8) imply that

$$\lim_{\Omega\to\infty} \overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) = \begin{cases} \exp\Big\{-V_1(\widetilde{\boldsymbol{x}})\Big\} \,/ \sum_{\boldsymbol{u}\in\mathcal{G}_0} \exp\Big\{-V_1(\boldsymbol{u})\Big\}, & \text{for } \widetilde{\boldsymbol{x}} \in \mathcal{G}_0 \\ 0, & \text{for } \widetilde{\boldsymbol{x}} \notin \mathcal{G}_0 \end{cases} \tag{2.5.10}$$

with $\mathcal{G}_0$ being the set of all *ground states* (global minima) of the potential energy landscape V_0. As a consequence, for sufficiently large Ω such that $V(\widetilde{\boldsymbol{x}};\Omega) \simeq V_0(\widetilde{\boldsymbol{x}}) + \Omega^{-1}V_1(\widetilde{\boldsymbol{x}})$, the probability of a ground state of V_0 is determined by the potential energy function V_1. Moreover, only the ground states of V_0 have a non-negligible probability to be observed as Ω becomes large because $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ decays to zero as $\Omega \to \infty$, for every $\widetilde{\boldsymbol{x}} \notin \mathcal{G}_0$. These results imply that the master equation (2.1.8) will asymptotically converge, in the thermodynamic limit, almost surely to a ground state of the potential energy function V_0, independently of the initial state. The particular ground state is chosen with probability determined by the values of the potential energy function V_1 over the ground states of V_0. On the other hand, the macroscopic equations (2.3.25) might reach a minimum of V_0, which may or may not be a ground state, depending on the initial condition.

If the macroscopic equations have a unique stable solution at steady-state that is independent of the initial condition, then V_0 will have only one (global) minimum. In this case, and as we mentioned before, the master equation (2.1.8) will converge almost surely to the same state in the thermodynamic limit. However, if V_0 contains more than one minimum, then the stationary solution of the master equation (2.1.8) may be different from the stationary solution predicted by

the corresponding macroscopic equations (2.3.25). As a consequence,

$$\lim_{\Omega\to\infty}\lim_{t\to\infty} p_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}};t) \neq \lim_{t\to\infty}\lim_{\Omega\to\infty} p_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}};t) \tag{2.5.11}$$

in general. This distinct difference between the stationary behavior of the master equation (left-hand side of inequality) and of the macroscopic equations (right-hand side of inequality) is known as *Keizer's paradox* [100–102].

At finite but sufficiently large system sizes Ω, the peaks of the stationary probability distribution $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ will correspond to minima of the potential energy landscape $V(\widetilde{\boldsymbol{x}};\Omega) \simeq V_0(\widetilde{\boldsymbol{x}}) + \Omega^{-1}V_1(\widetilde{\boldsymbol{x}})$. Moreover, if $\widetilde{\boldsymbol{x}}'$ is a (strict) local minimum of $V_0(\widetilde{\boldsymbol{x}}) + \Omega^{-1}V_1(\widetilde{\boldsymbol{x}})$, then

$$V_0(\widetilde{\boldsymbol{x}}) > V_0(\widetilde{\boldsymbol{x}}')\left[1 - \frac{1}{\Omega}\frac{V_1(\widetilde{\boldsymbol{x}}) - V_1(\widetilde{\boldsymbol{x}}')}{V_0(\widetilde{\boldsymbol{x}}')}\right], \quad \text{for every } \widetilde{\boldsymbol{x}} \in \mathcal{W}(\widetilde{\boldsymbol{x}}'), \tag{2.5.12}$$

where $\mathcal{W}(\widetilde{\boldsymbol{x}}')$ is a local neighborhood of $\widetilde{\boldsymbol{x}}'$ that does not contain $\widetilde{\boldsymbol{x}}'$ for which the inequality is satisfied. However, and for large enough Ω, such that

$$\frac{1}{\Omega}\left[\frac{V_1(\widetilde{\boldsymbol{x}}) - V_1(\widetilde{\boldsymbol{x}}')}{V_0(\widetilde{\boldsymbol{x}}')}\right] \simeq 0, \quad \text{for every } \widetilde{\boldsymbol{x}} \in \mathcal{W}(\widetilde{\boldsymbol{x}}'), \tag{2.5.13}$$

we approximately have $V_0(\widetilde{\boldsymbol{x}}') < V_0(\widetilde{\boldsymbol{x}})$, for every $\widetilde{\boldsymbol{x}} \in \mathcal{W}(\widetilde{\boldsymbol{x}}')$, and therefore $\widetilde{\boldsymbol{x}}'$ will approximately be a (strict) local minimum of the potential energy landscape V_0. Likewise, if $\widetilde{\boldsymbol{x}}'$ is a (strict) local minimum of V_0, then it will also be a (strict) local minimum of $V_0(\widetilde{\boldsymbol{x}}) + \Omega^{-1}V_1(\widetilde{\boldsymbol{x}})$, provided that Eq. (2.5.13) is satisfied. Hence, the minima of the potential energy landscape $V_0(\widetilde{\boldsymbol{x}}) + \Omega^{-1}V_1(\widetilde{\boldsymbol{x}})$ will correspond in this case to the stable stationary states of the macroscopic equations (2.3.25). As a consequence, the peaks of the stationary probability distribution $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ will correspond to stable stationary states of the macroscopic equations. For this reason, we refer to the peaks in $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ as *macroscopic* modes. Note however that there might be stable stationary states of the macroscopic equations that do not introduce peaks in the stationary probability distribution. To see this, recall that, in the thermodynamic limit as $\Omega \to \infty$, the peaks present in the stationary probability distribution are the ones associated only with the global minima of V_0.

At smaller values of Ω, the stationary probability distribution $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ will be given by Eqs. (2.5.2) and (2.5.3). The modes will now depend on the fluctuation size parameter Ω and

will be determined by the minima of the potential energy landscape $V(\widetilde{\boldsymbol{x}};\Omega)$. However, a state that minimizes the potential energy function V may not necessarily minimize V_0, in which case at least some modes of the probability distribution $\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})$ will not be predicted by the corresponding macroscopic equations. These modes are referred to as *noise-induced* modes, since they show up at small system sizes in which appreciable stochastic fluctuations may be present in the system due to "intrinsic noise." Recent literature has documented the presence of noise-induced modes in biochemical reaction networks and their importance in modeling system behavior not accounted for by their macroscopic counterparts [102–107].

Note finally that, if a Markovian reaction network is at a stable state $\widetilde{\boldsymbol{x}}_1^s$ at time t_0, then it may switch to another stable state $\widetilde{\boldsymbol{x}}_2^s$ at time $t_0 < t < \infty$ with probability $\Pr[\widetilde{\boldsymbol{X}}(t) = \widetilde{\boldsymbol{x}}_2^s \mid \widetilde{\boldsymbol{X}}(t_0) = \widetilde{\boldsymbol{x}}_1^s]$. However, $\lim_{\Omega\to\infty} \Pr[\widetilde{\boldsymbol{X}}(t) = \widetilde{\boldsymbol{x}}_2^s \mid \widetilde{\boldsymbol{X}}(t_0) = \widetilde{\boldsymbol{x}}_1^s] = \delta(\widetilde{\boldsymbol{x}}_2^s - \boldsymbol{\chi}(t))$, where $\delta(\cdot)$ is the Dirac delta function and $\boldsymbol{\chi}(t)$ is the solution of the macroscopic equations (2.3.25), initialized with $\widetilde{\boldsymbol{x}}_1^s$. Since $\widetilde{\boldsymbol{x}}_1^s$ is a minimum of the potential energy function V_0, the macroscopic system will be in state $\boldsymbol{\chi}(t) = \widetilde{\boldsymbol{x}}_1^s$ at time t. Hence, $\lim_{\Omega\to\infty} \Pr[\widetilde{\boldsymbol{X}}(t) = \widetilde{\boldsymbol{x}}_2^s \mid \widetilde{\boldsymbol{X}}(t_0) = \widetilde{\boldsymbol{x}}_1^s] = 0$. As a consequence, the probability of switching from a stable state to another stable state tends (in general exponentially) to zero as the system size increases to infinity. At finite system sizes Ω, switching among stable stationary states becomes possible, but the probability of switching is very small for large Ω; i.e., switching among stable stationary states are *rare* events [98, 107]. As a matter of fact, the waiting time for switching can be approximated by an exponential distribution [108] with rate parameter that tends to zero in the thermodynamic limit as $\Omega \to \infty$. Therefore, efficient switching between modes requires small system sizes and thus appreciable intrinsic noise.

2.6 Macroscopic (thermodynamic) behavior

We can view a Markovian reaction network as a thermodynamic system that absorbs energy, produces entropy, and dissipates heat [91, 102, 109–123]. This perspective can provide important

insights into functional properties of the network, such as robustness and stability, and can lead to a better understanding of the relationship between its mesoscopic (unobservable) and macroscopic (observable) behavior [93, 102, 112, 115, 117, 121, 124, 125].

In this section, we consider an *irreducible* Markovian reaction network comprised of $M/2$ pairs of *reversible* reactions $(2m-1, 2m)$, $m = 1, 2, \ldots, M/2$, where $2m-1$ is the forward reaction and $2m$ is the corresponding reverse reaction. This does not forbid us to consider irreversible reactions, since an irreversible reaction can be thought of as being reversible with negligible propensity in the reverse direction. As we mentioned in Section 2.4, the reaction network is characterized by a unique population probability distribution $p_{\boldsymbol{x}}(\boldsymbol{x}; t)$ that is analytic for all $t \geq 0$ and converges to a stationary distribution $\overline{p}_{\boldsymbol{x}}(\boldsymbol{x})$, which does not depend on the initial state $\boldsymbol{x}(0)$. By following our discussion in Section 2.5, we can define the *energy* of state $\boldsymbol{x}$ by

$$E(\boldsymbol{x}) := -\frac{1}{\Omega} \ln \overline{p}_{\boldsymbol{x}}(\boldsymbol{x}), \quad \text{for} \quad \boldsymbol{x} \in \mathcal{X}, \tag{2.6.1}$$

where $\Omega > 0$ is an appropriately chosen size parameter.

Our discussion in the following is purely mathematical in nature and can be applied to any physical or nonphysical Markovian reaction network. However, direct connection to thermodynamics can be made in certain physical systems, such as biochemical reaction networks, which may exchange matter, work, and heat through a well-defined boundary that separates the system from its surroundings [15]. In this case, we must take the size parameter Ω to be the inverse of $k_B T$, where k_B is the Boltzmann constant and T is the system temperature. Since the exact value of Ω is not important here, we set $\Omega = 1$ for simplicity.

By viewing a Markovian reaction network as a thermodynamic system, we can define three fundamental quantities: the internal energy, entropy, and Helmholtz free energy. The *internal energy* $U(t)$ is the average energy of the system at time t over all states, given by

$$U(t) := \sum_{\boldsymbol{x} \in \mathcal{X}} E(\boldsymbol{x}) p_{\boldsymbol{x}}(\boldsymbol{x}; t), \quad \text{for} \quad t \geq 0, \tag{2.6.2}$$

whereas, the entropy is defined by

$$S(t) := -\sum_{\boldsymbol{x}\in\mathcal{X}} p_{\boldsymbol{X}}(\boldsymbol{x};t)\ln p_{\boldsymbol{X}}(\boldsymbol{x};t), \quad \text{for} \quad t\geq 0. \tag{2.6.3}$$

Moreover, the Helmholtz free energy is given by

$$F(t) := U(t) - S(t) \ = \sum_{\boldsymbol{x}\in\mathcal{X}} p_{\boldsymbol{X}}(\boldsymbol{x};t)\ln\frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}, \quad \text{for} \quad t\geq 0. \tag{2.6.4}$$

The Helmholtz free energy measures the energy available in a thermodynamic system to do work under constant temperature and volume. Note that $F(t)$ coincides with the Kullback-Leibler distance of the probability distribution $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ from the steady-state probability distribution $\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})$ [recall Eq. (2.4.5)]. Therefore, the Helmholtz free energy provides a measure of how far a Markovian reaction network is from steady-state at time t. It turns out that $F(t)\geq 0$ and $dF(t)/dt\leq 0$, for every $t\geq 0$, with equality only at steady-state [91, 117, 126].

2.6.1 Balance equations

From Eq. (2.1.8) and Eq. (2.6.3), we can show (see Appendix A) the following *entropy balance* equation:

$$\frac{dS(t)}{dt} = \sigma(t) - h(t), \quad \text{for} \ \ t>0, \tag{2.6.5}$$

where

$$\sigma(t) = \frac{1}{2}\sum_{m=1}^{M/2}\sum_{\boldsymbol{x}\in\mathcal{X}}\Big[\,\rho_m^+(\boldsymbol{x};t)\mathcal{A}_m^+(\boldsymbol{x};t) + \rho_m^-(\boldsymbol{x};t)\mathcal{A}_m^-(\boldsymbol{x};t)\Big], \tag{2.6.6}$$

and

$$h(t) = \frac{1}{2}\sum_{m=1}^{M/2}\sum_{\boldsymbol{x}\in\mathcal{X}}\left\{\rho_m^+(\boldsymbol{x};t)\ln\left[\frac{\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})}{\pi_{2m}(\boldsymbol{x})}\right] + \rho_m^-(\boldsymbol{x};t)\ln\left[\frac{\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})}{\pi_{2m-1}(\boldsymbol{x})}\right]\right\}. \tag{2.6.7}$$

In these equations, $\rho_m^+(\boldsymbol{x};t)$ is the *net flux* of the m-th pair of reversible reactions reaching state $\boldsymbol{x}$ from state $\boldsymbol{x}-\boldsymbol{s}_{2m-1}$, given by $\rho_m^+(\boldsymbol{x};t) = \pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t) - \pi_{2m}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)$, whereas, $\rho_m^-(\boldsymbol{x};t)$ is the net flux of the same pair of reactions reaching state $\boldsymbol{x}$ from state $\boldsymbol{x}-\boldsymbol{s}_{2m}$, given by $\rho_m^-(\boldsymbol{x};t) = \pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t) - \pi_{2m-1}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)$ [note that $\boldsymbol{s}_{2m} = -\boldsymbol{s}_{2m-1}$].

Moreover,

$$\mathcal{A}_m^+(\boldsymbol{x};t) := \ln\left[\frac{\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)}{\pi_{2m}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)}\right]$$
$$\mathcal{A}_m^-(\boldsymbol{x};t) := \ln\left[\frac{\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)}{\pi_{2m-1}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)}\right] \tag{2.6.8}$$

are the *affinities* corresponding to the net fluxes $\rho_m^+(\boldsymbol{x};t)$ and $\rho_m^-(\boldsymbol{x};t)$, respectively. Note that $\rho_m^-(\boldsymbol{x};t) = -\rho_m^+(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)$ and $\mathcal{A}_m^-(\boldsymbol{x};t) = -\mathcal{A}_m^+(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)$, whereas,

$$\frac{\partial p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\partial t} = \sum_{m=1}^{M/2}\Big[\rho_m^+(\boldsymbol{x};t)+\rho_m^-(\boldsymbol{x};t)\Big], \quad t>0. \tag{2.6.9}$$

Therefore, $[\rho_m^+(\boldsymbol{x};t)+\rho_m^-(\boldsymbol{x};t)]dt$ quantifies the change [increase, when $\rho_m^+(\boldsymbol{x};t)+\rho_m^-(\boldsymbol{x};t)>0$, or decrease, when $\rho_m^+(\boldsymbol{x};t)+\rho_m^-(\boldsymbol{x};t)<0$] in the probability mass of the population process within the infinitesimally small time interval $[t, t+dt)$ due to the m-th pair of reversible reactions. These changes are driven by the affinities $\mathcal{A}_m^+(t)$ and $\mathcal{A}_m^-(t)$, which can be viewed as *thermodynamic forces* that move a Markovian reaction network away from the state of *thermodynamic equilibrium* (see Section 2.6.2), in which all net fluxes are zero.

Equation (2.6.5) provides an expression for the rate of entropy change in a Markovian reaction network. The term $\sigma(t)$ quantifies the rate of entropy production, whereas, the term $h(t)$ quantifies the rate of entropy loss due to heat dissipation. For this reason, $\sigma(t)$ and $h(t)$ are called the *entropy production rate* and the *heat dissipation rate*, respectively. On the other hand, Eq. (2.6.6) shows that $\sigma(t)$ is a sum of terms $1/2\sum_{\boldsymbol{x}\in\mathcal{X}}\big[\rho_m^+(\boldsymbol{x};t)\mathcal{A}_m^+(\boldsymbol{x};t)+\rho_m^-(\boldsymbol{x};t)\mathcal{A}_m^-(\boldsymbol{x};t)\big]$, each quantifying the contribution of a pair of reversible reactions to the net rate of entropy production. Similarly, Eq. (2.6.7) shows that $h(t)$ is a sum of terms $1/2\sum_{\boldsymbol{x}\in\mathcal{X}}\big\{\rho_m^+(\boldsymbol{x};t)\ln[\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})/\pi_{2m}(\boldsymbol{x})]+\rho_m^-(\boldsymbol{x};t)\ln[\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})/\pi_{2m-1}(\boldsymbol{x})]\big\}$, each quantifying the contribution of a pair of reversible reactions to the net rate of heat dissipation. Therefore, a reaction with non-zero net flux must produce entropy and dissipate heat.

By differentiating Eq. (2.6.4) with respect to t and by using Eq. (2.1.8), Eq. (2.6.6) and Eq. (2.6.8), we can derive (see Appendix A) the following balance equations for the Helmholtz free

energy and internal energy:

$$\frac{dF(t)}{dt} = f(t) - \sigma(t), \quad \text{for } t > 0, \tag{2.6.10}$$

and

$$\frac{dU(t)}{dt} = f(t) - h(t), \quad \text{for } t > 0, \tag{2.6.11}$$

where

$$f(t) := \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x} \in \mathcal{X}} \Big[\rho_m^+(\boldsymbol{x}; t) \bar{\mathcal{A}}_m^+(\boldsymbol{x}) + \rho_m^-(\boldsymbol{x}; t) \bar{\mathcal{A}}_m^-(\boldsymbol{x}) \Big], \tag{2.6.12}$$

with $\bar{\mathcal{A}}_m^+(\boldsymbol{x})$ and $\bar{\mathcal{A}}_m^-(\boldsymbol{x})$ being the affinities of the m-th pair of reversible reactions at steady-state; i.e., $\bar{\mathcal{A}}_m^+(\boldsymbol{x}) := \lim_{t \to \infty} \mathcal{A}_m^+(\boldsymbol{x}; t)$ and $\bar{\mathcal{A}}_m^-(\boldsymbol{x}) := \lim_{t \to \infty} \mathcal{A}_m^-(\boldsymbol{x}; t)$. Equation (2.6.10) quantifies the change in Helmholtz free energy due to the Markovian reaction network being away from thermodynamic equilibrium at steady-state [quantified by the first term on the right-hand-side of Eq. (2.6.10)] or reduction in Helmholtz free energy due to entropy production [quantified by the second term on the right-hand-side of Eq. (2.6.10)]. The term $f(t)$ quantifies the rate of energy (i.e., power) supplied to the Markovian reaction network in order to keep it away from thermodynamic equilibrium. For this reason, we refer to $f(t)$ as the *"motive" power*. This quantity is also known in the literature as the rate of "housekeeping" heat [110, 114, 116, 118, 120]. However, we prefer to call $f(t)$ the "motive" power, since it represents the energy flow per unit time required to keep the Markovian reaction network away from thermodynamic equilibrium. It turns out that $0 \leq f(t) \leq \sigma(t)$, for every $t \geq 0$. We can show the first inequality by using the fact that the right-hand side of the master equation (2.1.8) is zero at steady-state and that $\ln x \leq x - 1$, for $x > 0$ (see [116]). The second inequality is due to Eq. (2.6.10) and the fact that $dF(t)/dt \leq 0$. Note that $f(t)$ is a sum of terms $1/2 \sum_{\boldsymbol{x} \in \mathcal{X}} \big[\rho_m^+(\boldsymbol{x}; t) \bar{\mathcal{A}}_m^+(\boldsymbol{x}) + \rho_m^-(\boldsymbol{x}; t) \bar{\mathcal{A}}_m^-(\boldsymbol{x}) \big]$, each term quantifying the contribution of a pair of reversible reactions to the net "motive" power. Therefore, a reaction with non-zero (forward or reverse) flux and corresponding non-zero affinity at steady-state will supply motive power to the Markovian reaction network.

Equation (2.6.10) shows that reactions in a Markovian reaction network can increase the

Helmholtz free energy by adding "motive" energy to the system, whereas, they can reduce the Helmholtz free energy due to entropy production. Moreover,

$$\sigma(t) = f(t) + \left|\frac{dF(t)}{dt}\right|, \quad \text{for } t > 0, \tag{2.6.13}$$

which implies that entropy production comes from two sources: from supplying motive power $f(t)$ to sustain the reaction network away from thermodynamic equilibrium and from a spontaneous change $|dF(t)/dt|$ in Helmholtz free energy due to relaxation towards the steady-state [127]. On the other hand, Eq. (2.6.11) expresses the first-law of thermodynamics (energy conservation): a change $\Delta U(t) = U(t+dt) - U(t)$ in internal energy within an infinitesimal time interval $[t, t+dt)$ must equal the amount of motive energy $f(t)dt$ added to the system minus the dissipated heat $h(t)dt$. From Eq. (2.6.6), note that $\sigma(t) \geq 0$, for every $t \geq 0$, with equality if and only if $\mathcal{A}_m^+(t) = \mathcal{A}_m^-(t) = 0$, for every $m = 1, 2, \ldots, M/2$, which is a direct consequence of the fact that $(x_1 - x_2)\ln(x_1/x_2) \geq 0$, for any values of x_1 and x_2, with equality if and only if $x_1 = x_2$. This result is in agreement with the second law of thermodynamics, which postulates that the rate of entropy production must always be nonnegative. Finally, Eq. (2.6.5) and Eq. (2.6.10) imply that

$$0 \leq \overline{\sigma} = \overline{h} = \overline{f}, \tag{2.6.14}$$

where $\overline{\sigma} := \lim_{t\to\infty} \sigma(t)$, and similarly for $\overline{h}$ and $\overline{f}$. This result implies that, at steady-state, the amount of motive power supplied to the system must be equal to the rate of heat dissipation, in agreement with the first law of thermodynamics. Moreover, the rate of heat dissipation must be equal to the rate of entropy production. It also implies that the steady-state entropy production, heat dissipation and motive power must all be nonnegative, in agreement with the second law of thermodynamics.

2.6.2 Thermodynamic equilibrium

A Markovian reaction network reaches thermodynamic equilibrium at steady-state if and only if $\bar{\mathcal{A}}_m^+ = \bar{\mathcal{A}}_m^- = 0$, for every $m = 1, 2, \ldots, M/2$, which is equivalent to the following *detailed*

balance equations:

$$\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1}) = \pi_{2m}(\boldsymbol{x})\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}) \tag{2.6.15}$$

$$\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1}) = \pi_{2m-1}(\boldsymbol{x})\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}), \tag{2.6.16}$$

for every $m = 1, 2, \ldots, M/2$, $\boldsymbol{x} \in \mathcal{X}$. In this case, $f(t) = 0$, for every $t \geq 0$, which implies that

$$\frac{dU(t)}{dt} = -h(t) \quad \text{and} \quad \frac{dF(t)}{dt} = -\sigma(t), \quad \text{for } t > 0. \tag{2.6.17}$$

Moreover, Eq. (2.6.14) results in $\overline{\sigma} = \overline{h} = \overline{f} = 0$, which shows that a Markovian reaction network that reaches thermodynamic equilibrium at steady-state will not produce entropy or dissipate heat. It turns out that a Markovian reaction network must be *reversible* at thermodynamic equilibrium, which means that the stationary behavior of the population process will be indistinguishable if the direction of time is reversed. This behavior may not be desirable, since many Markovian reaction systems (e.g., biochemical reaction networks) are irreversible with respect to time. As a matter of fact, entropy production, heat dissipation, and irreversibility with respect to time are three properties necessary for the formation of order in physical systems [109]. As a consequence, a useful Markovian reaction network must not reach thermodynamic equilibrium in most cases of interest. We can make sure that this is the case by including nonreversible reactions that transfer mass between the system and its surroundings, thus breaking detailed balance.

Despite the aforementioned drawbacks, Markovian reaction networks that reach thermodynamic equilibrium have been extensively used to model population dynamics. For this type of networks we can use (at least in principle) a simple iterative procedure to calculate the steady-state probability distribution. This is possible because *any* state $\boldsymbol{x} \in \mathcal{X}$ can be reached from a given state $\boldsymbol{x}_0 \in \mathcal{X}$ through at least one ordered chain of reactions $(m_1, m_2, \ldots, m_L)$. In this case, detailed balance implies that [128]

$$\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}) = \overline{p}_{\boldsymbol{X}}(\boldsymbol{x}_0) \prod_{l=1}^{L} \frac{\pi_{m_l}(\boldsymbol{x}_0 + \sum_{l'=1}^{l-1} \boldsymbol{s}_{m_{l'}})}{\pi_{m_l^*}(\boldsymbol{x}_0 + \sum_{l'=1}^{l} \boldsymbol{s}_{m_{l'}})}, \tag{2.6.18}$$

for every $\boldsymbol{x} \neq \boldsymbol{x}_0$, where m_l^* is the index of the opposite reaction to reaction m_l (i.e., $m_l^* = 2m$, if $m_l = 2m-1$, and $m_l^* = 2m-1$, if $m_l = 2m$). After this procedure is completed for all $\boldsymbol{x} \in \mathcal{X}$, we can calculate $\overline{p}_{\boldsymbol{x}}(\boldsymbol{x}_0)$ in Eq. (2.6.18) by setting the sum of all probabilities $\overline{p}_{\boldsymbol{x}}(\boldsymbol{x})$ equal to $1 - \overline{p}_{\boldsymbol{x}}(\boldsymbol{x}_0)$.

Chapter 3

Numerically Solving the Master Equation: Implicit Euler Method

In the following[1], we present a novel numerical algorithm for solving the master equation. We demonstrate that this algorithm is particularly useful in a wide class of processes on networks studied by epidemiologists.

The processes by which disease spreads in a population of individuals are inherently stochastic. The master equation has proven to be a useful tool for modeling such processes. Unfortunately, as we mentioned in the previous chapter, solving the master equation analytically is possible only in limited cases (e.g., when the model is linear), and thus numerical procedures or approximation methods must be employed. Available techniques, such as the LNA method, may fail to provide reliable solutions, whereas current numerical routines can induce unreasonable computational burden.

In this chapter, we propose a new numerical technique for solving the master equation. Our method is based on a more informative stochastic process than the population process commonly used in the literature. By exploiting the structure of the master equation governing this process, we

[1]The material in this chapter is reprinted with permission from "Numerical Integration of the Master Equation in Some Models of Stochastic Epidemiology", by Garrett Jenkinson and John Goutsias, *PLoS One*, vol. 7 issue 5, number e36160. Copyright 2012, Jenkinson & Goutsias.

develop a novel technique for calculating the *exact* solution—up to desired computational precision—of the master equation in certain models with inherent structure that is common in stochastic epidemiology. We demonstrate the potential of our method by solving the master equation associated with the stochastic SIR epidemic model.

3.1 Motiviation

Stochasticity can play an important role when studying a disease that spreads through a population of individuals [17, 92, 129]. A common approach to modeling this problem is by means of a Markov process, whose probability distribution satisfies a master equation. Solving this equation analytically however is not in general possible and Monte Carlo sampling, based on the Gillespie algorithm [45], is often used to accomplish this goal. Unfortunately, accurate evaluation of the probability distribution of a Markov process requires a prohibitively large number of Monte Carlo samples for most systems of interest. As a consequence, Monte Carlo sampling is mostly used to estimate statistical summaries of the underlying stochastic population dynamics, such as means and variances.

To evaluate the solution of the master equation, a number of approximation techniques have been proposed in the literature, such as the LNA method [130]. While this approximation may work well in certain circumstances, it often fails when the underlying assumptions are not satisfied. The LNA method can only produce a normal approximation to the solution of the master equation. Therefore, if the probability distribution of the population process is bimodal, then this method will produce erroneous results.

Some effort has recently shifted away from Monte Carlo sampling and approximation techniques and has focused on exploiting the linear structure of the master equation associated with the population process. This results in a numerical solution to the master equation through matrix exponentiation; e.g., see [30, 32, 33, 92, 131]. A popular technique along these lines employs a Krylov

subspace approximation (KSA) method [32, 33] that dramatically reduces the size of matrix exponentiation and results in an attractive iterative algorithm for solving the master equation. However, the KSA technique is based on several approximations, whose cumulative effect may appreciably affect the method's accuracy, numerical stability, and computational efficiency.

There are two main issues that can affect performance of the KSA method. One is choosing the dimension of the approximating Krylov subspace used. If the dimension is chosen too small, the method may produce an inaccurate solution to the master equation, whereas, a value that is too large can result in an appreciable decrease of computational efficiency. Unfortunately, there is no rigorous way to optimally determine an appropriate value for this parameter, which is chosen manually, even in advanced implementations such as Expokit [32]. Another issue is the fact that, at each step, the KSA method may not necessarily produce a probability vector (i.e., a vector composed of nonnegative elements that sum to one). This problem can be addressed by using a sufficiently small step-size, but this may seriously affect the method's computational efficiency. In practice, the KSA method is equipped with a heuristic step that zeros-out all negative values and re-normalizes the positive values so that they sum to one. This step however introduces its own errors, which may affect the quality of the approximation in an manner that is not easy to predict.

Instead of using the population process, we can describe the stochastic spread of a disease by a more informative stochastic process known as the degree-of-advancement (DA). Exploiting the structure of the master equation governing this process results in a novel numerical algorithm for calculating the *exact* solution of the master equation, which we refer to as the implicit Euler (IE) method. This technique enjoys several advantages over the KSA method: its global error is of first-order with respect to the step-size, it is numerically stable regardless of the step-size used, and always produces a solution whose elements are nonnegative and sum to one. As we will discuss in this chapter, the IE method shows great promise for solving certain problems in stochastic epidemiology in which the state-space associated with the DA process is reasonably sized. It is not however meant to replace the KSA method, which is still the best numerical method available for solving the master

equation in problems where implementation of the IE method is not computationally attractive or possible. To illustrate the potential of the proposed IE method, we calculate the *exact* solution of the master equation associated with the stochastic SIR epidemic model and use this solution to study some important properties of this model.

3.2 Methods

3.2.1 Disease dynamics

As discussed in Section 2.2.2, the classical SIR epidemic model (without births, deaths, or imports of disease) is one of the simplest models in epidemiology. Here, each individual in a population is either susceptible to a disease, infected, or recovered. If we denote by S, I, and R the susceptible, infected and recovered individuals, respectively, and by $S(t)$, $I(t)$ and $R(t)$ their corresponding (and possibly random) population numbers, we can characterize the state of the SIR model at time t by using the 3×1 vector $[S(t)\ I(t)\ R(t)]^T$, where T denotes vector transpose. The state depends on time due to the (possibly random) occurrences of the following two reactions:

$$\mathrm{S} + \mathrm{I} \to 2\mathrm{I} \quad \text{and} \quad \mathrm{I} \to \mathrm{R}, \tag{3.2.1}$$

which model infection of a susceptible individual (first reaction) as well as recovery of an infected individual (second reaction).

We can model a complex epidemiological system in more general terms by using the general reaction form given by Eq. (2.1.1). This model congregates individuals into N different groups, $X_1, X_2, \ldots, X_N$, which interact through M coupled reactions. For example, in the aforementioned SIR model, we may set $X_1 = S$, $X_2 = I$, $X_3 = R$, resulting in $\nu_{11} = \nu_{21} = \nu_{22} = \nu'_{32} = 1$, $\nu'_{21} = 2$, with the remaining coefficients being zero.

The usual way to characterize an epidemiological system is by means of the $N \times 1$ random vector $\boldsymbol{X}(t)$ with elements $X_n(t), n \in \mathcal{N}$, where $X_n(t)$ denotes the population of the n-th group of

individuals present in the system at time $t \geq 0$. By convention, we set $\boldsymbol{X}(0) = \boldsymbol{x}(0)$, for some known value $\boldsymbol{x}(0)$ (i.e., we assume that we know the initial population numbers at time $t = 0$). We refer to the multivariate stochastic process $\{\boldsymbol{X}(t), t > 0\}$ as the *population process*.

Let $Z_m(t)$ be the (possibly random) number of times that the m-th reaction occurs during the time interval $[0, t)$. Then, $\{Z_m(t), t > 0\}$ is a counting process, known as the *degree of advancement* (DA) of the m-th reaction [1]. We set $Z_m(0) := 0$ and refer to the multivariate stochastic process $\{\boldsymbol{Z}(t), t \geq 0\}$ as the *DA process*. Note that according to Eq. (2.1.4) we have $\boldsymbol{X}(t) = \boldsymbol{x}(0) + \mathbb{S}\boldsymbol{Z}(t)$. Thus, given an initial population vector $\boldsymbol{x}(0)$, Eq. (2.1.4) allows us to uniquely determine the population process $\boldsymbol{X}(t)$ from the DA process $\boldsymbol{Z}(t)$. However, we cannot in general determine the DA process from the population process. This can only be done when the nullity of $\mathbb{S}$ is zero, in which case $\boldsymbol{Z}(t) = (\mathbb{S}^T\mathbb{S})^{-1}\mathbb{S}^T[\boldsymbol{X}(t) - \boldsymbol{x}(0)]$. As a consequence, the DA process is more informative than the population process. The DA process' probability mass function $p_{\boldsymbol{Z}}(\boldsymbol{z}; t)$ is governed by the master equation (2.1.6). We can use the solution $p_{\boldsymbol{Z}}(\boldsymbol{z}; t)$ of the previous master equation to calculate the probability mass function $p_{\boldsymbol{X}}(\boldsymbol{x}; t)$ of the population process according to Eq. (2.1.9)

3.2.2 Exploiting structure

Most available algorithms for solving the master equation focus on the population process instead of the DA process. It turns out that, by using the DA process, we may reap some benefits that can lead to a simple numerical solver for the general master equation (2.1.6).

In the following, we assume that statistical analysis of an epidemiological model of interest is limited within a finite time interval $\mathcal{T} := [0, t_{\max}]$, where the maximum time $t_{\max}$ is such that the the DA process is almost surely contained within $\mathcal{Z}$, which is an M-dimensional discrete and finite sample space, i.e.

$$\sum_{\boldsymbol{z} \in \mathcal{Z}} p_{\boldsymbol{Z}}(\boldsymbol{z}; t) = 1, \quad \text{for every } t \in \mathcal{T}. \tag{3.2.2}$$

We index the elements in $\mathcal{Z}$ by $\boldsymbol{z}_k$, $k = 1, 2, \ldots, K$, where K is the cardinality of $\mathcal{Z}$ (i.e., the total

number of elements in $\mathcal{Z}$). We can then define the $K \times 1$ vector $\boldsymbol{q}(t)$ with elements $q_k(t) = p_{\boldsymbol{Z}}(\boldsymbol{z}_k; t)$, for $k = 1, 2, \ldots, K$. Clearly, $\boldsymbol{q}(t)$ specifies the probability mass function $p_{\boldsymbol{Z}}(\boldsymbol{z}; t)$. It can be seen from Eq. (2.1.6) that $\boldsymbol{q}(t)$ can be calculated by solving the following system of K *linear* ordinary differential equations (ODEs):

$$\frac{d\boldsymbol{q}(t)}{dt} = \mathbb{Q}\boldsymbol{q}(t), \quad t \in \mathcal{T}, \tag{3.2.3}$$

where $\mathbb{Q}$ is a $K \times K$ matrix that can be directly constructed from the master equation. In the theory of Markov processes, $\mathbb{Q}$ is known as the *generator matrix*. Note that the k-th column of $\mathbb{Q}$ contains zeros in most places except for the k-th element that takes value $-\sum_{m \in \mathcal{M}} \alpha_m(\boldsymbol{z}_k) \leq 0$ and M off-diagonal elements that take values $\alpha_m(\boldsymbol{z}_k) \geq 0$, $m \in \mathcal{M}$. Therefore, the elements of each column of $\mathbb{Q}$ add to zero; see Appendix B for an example. Finally, Eq. (3.2.3) is initialized by a vector $\boldsymbol{q}(0)$ whose first element equals 1 (assuming that $\boldsymbol{z}_1 = \boldsymbol{0}$), whereas, the remaining elements are all zero.

The main advantage of using the DA process $\boldsymbol{Z}(t)$ is that, under an appropriate ordering of the elements in $\mathcal{Z}$, the generator matrix $\mathbb{Q}$ will be *lower triangular*, a result that is not true when employing the population process $\boldsymbol{X}(t)$. We will shortly demonstrate that this can result in substantial simplification of the numerical algorithm used to solve Eq. (3.2.3).

To obtain a matrix $\mathbb{Q}$ that is lower triangular, we must order the points $\boldsymbol{z}_k$ in the sample space $\mathcal{Z}$ *lexicographically*, such that $\boldsymbol{z}_k \prec \boldsymbol{z}_{k+1}$, for $k = 1, 2, \ldots, K-1$, where $\prec$ denotes that one variable is lexicographically smaller than another [e.g., $(z_1, z_2) \prec (z_1', z_2')$ if and only if $z_1 < z_1'$ or $z_1 = z_1'$ and $z_2 < z_2'$]. Because a reaction can only *increase* (by one) the value of a single element of $\boldsymbol{z}$, it is not possible for probability mass to be transferred from $\boldsymbol{z}_{k'}$ to $\boldsymbol{z}_k$ when $\boldsymbol{z}_k \prec \boldsymbol{z}_{k'}$. Such monotonic transfer of probability does not generally occur when the population process $\boldsymbol{X}(t)$ is used. Therefore, when the points $\boldsymbol{z}_k$, $k = 1, 2, \ldots, K$, in $\mathcal{Z}$ are ordered lexicographically, the (k, k') element of matrix $\mathbb{Q}$ will be zero when $k' > k$ and, therefore, $\mathbb{Q}$ will be lower triangular. See Appendix B for an illustration.

3.2.3 Numerical solver

We now proceed by exploiting the three key structural characteristics of matrix $\mathbb{Q}$: its stability, triangularity, and sparsity. We have noted that the diagonal elements of $\mathbb{Q}$ are non-positive. However, since $\mathbb{Q}$ is triangular, its diagonal elements will be the eigenvalues of $\mathbb{Q}$. Thus, the linear constant coefficient system of ODEs given by Eq. (3.2.3) is *stable*, ensuring the efficacy of implicit ODE solvers [132]. As a consequence, we can use the implicit Euler method to estimate $\boldsymbol{q}(t)$ at discrete time points $t_j := j\tau$, $j = 1, 2, \ldots$, for a given time step τ. Then, given an estimate $\widehat{\boldsymbol{q}}(t_{j-1})$ of $\boldsymbol{q}(t_{j-1})$, we can obtain an estimate $\widehat{\boldsymbol{q}}(t_j)$ of $\boldsymbol{q}(t_j)$ by solving the following system of linear equations:

$$(\mathbb{I} - \tau\mathbb{Q})\,\widehat{\boldsymbol{q}}(t_j) = \widehat{\boldsymbol{q}}(t_{j-1}), \tag{3.2.4}$$

where $\mathbb{I}$ is the $K \times K$ identity matrix. In Appendix B, we show that solving the previous system is always possible, for any $\tau > 0$, due to the invertibility of matrix $\mathbb{I} - \tau\mathbb{Q}$. By initializing the computation with $\widehat{\boldsymbol{q}}(0) = \boldsymbol{q}(0)$, we can recursively calculate the values of the probability mass function $p_{\boldsymbol{Z}}(\boldsymbol{z}; t)$ of the DA process at the discrete time points t_j, $j = 1, 2, \ldots$. We also show in Appendix B that the previous procedure always returns a probability vector for any step-size $\tau \geq 0$. Moreover, we demonstrate that the resulting method is a first-order solver, since the *global* error $||\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j)||_1$ is of $\mathcal{O}(\tau)$ (i.e., the global error is proportional to the step-size τ). Finally, since the implicit Euler method is always stable for *any* choice of τ [132], the errors from previous iterations will not be amplified in later stages, regardless of the step-size used. Therefore, a desired error can be achieved by simply reducing the value of the step-size τ. We refer to the resulting technique for solving the master equation based on Eq. (3.2.4) as the *implicit Euler* (IE) method.

In general, solving Eq. (3.2.4) would require $\mathcal{O}(K^3)$ computations, where K is the cardinality of the sample space $\mathcal{Z}$, which will be prohibitive. However, since $\mathbb{Q}$ is a *triangular* matrix, we can use forward substitution whose cost is usually of $\mathcal{O}(K^2)$. But since $\mathbb{Q}$ is a *sparse* matrix, with each column having only $M + 1$ non-zero elements, forward substitution can be done at a cost

of $\mathcal{O}(MK)$ [133], where M is the number of reactions. In addition, calculating the probability mass function at time t_j requires storage of $\mathcal{O}(MK)$ nonzero numbers. In particular, we need to store MK nonzero elements of matrix $\mathbb{I} - \tau\mathbb{Q}$ as well as $2(K-1)$ elements of vectors $\widehat{\boldsymbol{q}}(t_j)$ and $\widehat{\boldsymbol{q}}(t_{j-1})$ [note that the elements of each column of matrix $\mathbb{I} - \tau\mathbb{Q}$ and the elements of each of the two vectors $\widehat{\boldsymbol{q}}(t_j)$ and $\widehat{\boldsymbol{q}}(t_{j-1})$ sum to one]. Since $K \gg M$, the computational and memory requirements of the IE method will be $\mathcal{O}(K)$, which grow *linearly* in terms of K.

3.2.4 Practical considerations

In general, the computational and memory requirements of matrix exponentiation grow *quadratically* in terms of the cardinality L of the sample space $\mathcal{X}$, and can quickly become prohibitive for large values of L. The KSA method however can greatly reduce this expense to $\mathcal{O}(L_0(M+L_0)L)$ computations and $\mathcal{O}((M+L_0)L)$ memory locations, where L_0 is the dimension of the approximating Krylov subspace used and M is the number of reactions (see Appendix B). Thus, the relative efficiency of the IE method, which requires $\mathcal{O}(MK)$ computation and storage cost, to the KSA approach will depend on the relative values of the cardinalities K and L of the sample spaces $\mathcal{Z}$ and $\mathcal{X}$, respectively.

As we mentioned before, if the nullity of the net stoichiometry matrix $\mathbb{S}$ is zero, then there is a one-to-one correspondence between $\boldsymbol{x} = \boldsymbol{x}(0) + \mathbb{S}\boldsymbol{z}$ and $\boldsymbol{z}$. As a consequence of Eq. (2.1.9), the cardinalities of $\mathcal{X}$ and $\mathcal{Z}$ will be the same, in which case $K = L$. Under these circumstances, the IE method will outperform the KSA method. This is a consequence of the fact that $MK = ML < L_0(M + L_0)L$ and $MK = ML < (M + L_0)L$ in this case. We can easily verify that, for the simple SI model ($\mathrm{S} + \mathrm{I} \to 2\mathrm{I}$), the SIR epidemic model characterized by Eq. (3.2.1), and the SEIR model ($\mathrm{S} + \mathrm{I} \to \mathrm{E} + \mathrm{I}$, $\mathrm{E} \to \mathrm{I}$, $\mathrm{I} \to \mathrm{R}$, where E denotes a group of individuals exposed to disease but not yet infectious), the nullity of $\mathbb{S}$ is indeed zero and, therefore, the IE method will be superior to the KSA method.

In general, the IE method will be computationally superior to the KSA method, provided

that the cardinality of the sample space $\mathcal{Z}$ is not appreciably larger than $L_0(M+L_0)/M$ times the cardinality of the sample space $\mathcal{X}$ [or not much larger than $(M+L_0)/M$ times the cardinality of the sample space $\mathcal{X}$, if we also consider memory requirements]. Of course, in situations where the nullity of $\mathbb{S}$ is large, the sample space $\mathcal{Z}$ can become appreciably larger than $\mathcal{X}$, in which case the KSA method will be more preferable. Note that there are cases in which $\mathcal{Z}$ and $\mathcal{X}$ can become infinite (e.g., suppose an influx of people at some constant rate $\emptyset \to X_n$, in which case both sample spaces will be unbounded). In these situations, the use of a finite state projection approach [30] is required to reduce the sample spaces, and the relative efficiency of the two methods will depend on the sizes of the resulting subspaces.

For a given step-size τ, the IE method described so far generates a sequence of probability vectors $\widehat{\boldsymbol{q}}(t_j)$, $j = 1, 2, \ldots$. Assuming that the true solution $\boldsymbol{q}(t_{j-1})$ is known at time t_{j-1}, we can show (see Appendix B) that the *local* error $\|\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j \mid t_{j-1})\|_1$ is of $\mathcal{O}(\tau^2)$, where $\widehat{\boldsymbol{q}}(t_j \mid t_{j-1})$ is the approximation of $\boldsymbol{q}(t_j)$ obtained by the IE method for a given value of $\boldsymbol{q}(t_{j-1})$. We can further improve this result by employing a powerful computational tool known as Richardson extrapolation [134].

We show in Appendix B that, if $\widehat{\boldsymbol{q}}_\tau(t_j \mid t_{j-1})$ and $\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})$ are the approximations of $\boldsymbol{q}(t_j)$ obtained from $\boldsymbol{q}(t_{j-1})$ by the IE method with step-sizes τ and $\tau/2$, respectively, then $\widehat{\boldsymbol{q}}_*(t_j \mid t_{j-1}) := 2\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) - \widehat{\boldsymbol{q}}_\tau(t_j \mid t_{j-1})$ also approximates $\boldsymbol{q}(t_j)$, but with a local error of $\mathcal{O}(\tau^3)$. We therefore expect that $\widehat{\boldsymbol{q}}_*(t_j \mid t_{j-1})$ is a better approximation to $\boldsymbol{q}(t_j)$ than $\widehat{\boldsymbol{q}}_\tau(t_j \mid t_{j-1})$ [or even $\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})$; see Appendix B] for a sufficiently small step-size τ. This suggests a valuable modification to the IE method that can be used to approximate the solution of the master equation better than the original technique. The modification combines two runs of the IE method, with time steps τ and $\tau/2$, and produces a solution $\widehat{\boldsymbol{q}}_*(t_j)$, given by

$$\widehat{\boldsymbol{q}}_*(t_j) = \begin{cases} 2\widehat{\boldsymbol{q}}_{\tau/2}(t_j) - \widehat{\boldsymbol{q}}_\tau(t_j), & \text{if } [2\widehat{\boldsymbol{q}}_{\tau/2}(t_j) - \widehat{\boldsymbol{q}}_\tau(t_j)]_{\min} \geq 0 \\ \widehat{\boldsymbol{q}}_{\tau/2}(t_j), & \text{otherwise,} \end{cases} \tag{3.2.5}$$

where $[\boldsymbol{x}]_{\min}$ denotes the minimum value of the elements of vector $\boldsymbol{x}$. In this case, $\widehat{\boldsymbol{q}}_*(t_j)$ is given

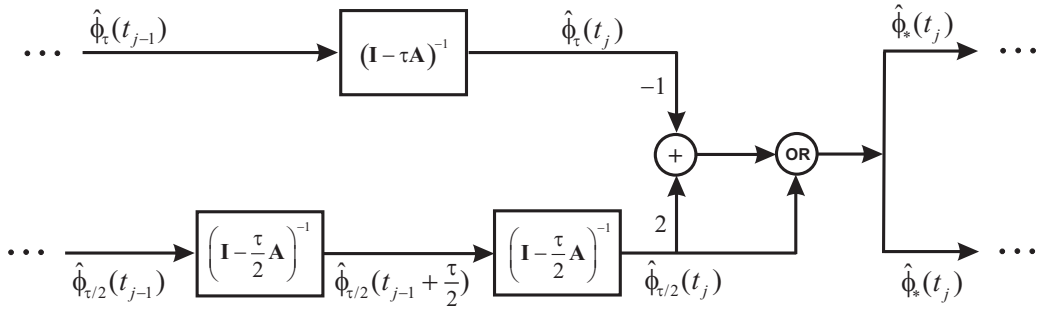

Figure 3.1: One step of the RIE method for solving the master equation. The upper branch implements the standard IE method with step-size τ, whereas, the lower branch implements the IE method with step-size $\tau/2$. "OR" implements Eq. (3.2.5).

by the "improved" vector $2\widehat{\boldsymbol{q}}_{\tau/2}(t_j) - \widehat{\boldsymbol{q}}_\tau(t_j)$ only when all elements of that vector are nonnegative. Otherwise, $\widehat{\boldsymbol{q}}_*(t_j)$ is given by the vector $\widehat{\boldsymbol{q}}_{\tau/2}(t_j)$ calculated by the IE method with the smaller step-size $\tau/2$. This assures that $\widehat{\boldsymbol{q}}_*(t_j)$ is always a probability vector. We will be referring to the resulting technique as the Richardson-based implicit Euler (RIE) method. We illustrate one step of this method in Fig. 3.1.

Many ODE solvers, including the KSA method, adjust the step-size at each iteration to assure that the local error ERR is less than a pre-specified error tolerance TOL while minimizing the computational effort required to accomplish this goal. We can also modify the RIE method to accommodate variable step-sizes. By following our analysis in Appendix B, we can approximately calculate the local error ERR_j at step j by

$$\mathrm{ERR}_j = 1.1 \times \|\widehat{\boldsymbol{q}}_{\tau/2}(t_j) - \widehat{\boldsymbol{q}}_\tau(t_j)\|_1, \tag{3.2.6}$$

where we use a factor of 1.1 to compensate for the possibility that the true (but unknown) local error is larger (by 10%) than the actual error calculated by $\|\widehat{\boldsymbol{q}}_{\tau/2}(t_j) - \widehat{\boldsymbol{q}}_\tau(t_j)\|_1$. If $\mathrm{ERR}_j < \mathrm{TOL}$, then we consider the step successful and increase the step-size from τ to τ^*, where

$$\tau^* = \tau\sqrt{\frac{\mathrm{TOL}}{\mathrm{ERR}_j}} = 0.95\,\tau\sqrt{\frac{\mathrm{TOL}}{\|\widehat{\boldsymbol{q}}_{\tau/2}(t_j) - \widehat{\boldsymbol{q}}_\tau(t_j)\|_1}}\,. \tag{3.2.7}$$

However, if $\mathrm{ERR}_j > \mathrm{TOL}$, then the step is unsuccessful. In this case, we decrease the step-size from τ to τ^* by using Eq. (3.2.7) and redo the RIE step.

Finally, we note that some users might be concerned with precision loss in the forward substitution step of the IE and RIE methods. The standard numerical technique of *iterative improvement* could be employed to protect against such precision loss [132], with moderate additional computational burden. However, we show in Appendix B that the matrix being inverted $(\mathbb{I} - \tau\mathbb{Q})$ is never singular, and it is readily apparent that for small τ this matrix is far from being singular. We thus suggest that reducing τ may be a preferable method of combating precision loss, since the step size also tightly regulates the global error as shown in Appendix B. In the following example, we did not perform iterative improvement and the results indicate that any precision loss was negligible despite the large dynamic range of probabilities involved in the solution.

3.3 Results

To demonstrate the efficacy of our method, we tackle the problem of modeling a well-documented 1978 influenza epidemic in an English boarding school [135]. A deterministic SIR model was originally developed to analyze these data [136]. Subsequently, the model was extended to the stochastic case and approximately solved using the LNA method [17]. In the following, we compute the *exact* solution of the underlying master equation. To the best of our knowledge, it is the first time an exact algorithm has been employed for this problem.

There are three classes of individuals, S, I and R, representing $Q = 763$ susceptible, infected and recovered pupils. Spreading of the epidemic is governed by the reactions in Eq. (3.2.1) with propensity functions

$$\pi_1(S(t), I(t), R(t)) = k_1 S(t) I(t) \quad \text{and} \quad \pi_2(S(t), I(t), R(t)) = k_2 I(t), \tag{3.3.8}$$

where $k_1 = 0.00218/\mathrm{day}$ and $k_2 = 0.44036/\mathrm{day}$ are the rate constants of infection and recovery,

respectively [17]. The initial conditions are given by

$$S(0) = 762, \quad I(0) = 1, \quad R(0) = 0, \tag{3.3.9}$$

reflecting the fact that only one pupil is infected at the start of the epidemic. We take the sample space $\mathcal{Z}$ to be the rectangular region in the $\boldsymbol{z}$ plane that begins at $(0, 0)$ and extends to include the maximal point $(762, 763)$. This is due to the fact that the first reaction can occur at most 762 times, after which all pupils will have been infected, whereas, the second reaction can occur at most 763 times, after which all pupils will have recovered from the infection. As a consequence, the sample space $\mathcal{Z}$ contains $K = 763 \times 764 = 582{,}932$ points.

Numerically solving the master equation over a period of 25 days by means of the KSA method using Expokit [32] took 72 minutes of CPU time on a 2.20 GHz Intel Mobile Core 2 Duo T7500 processor running Matlab 7.7. The resulting solution produces an L2 error $||\boldsymbol{p}(25) - \widehat{\boldsymbol{p}}(25)||_2 = 1.48 \times 10^{-3}$, where $\boldsymbol{p}$ is a solution of the master equation obtained by a stringent run of Expokit (see Appendix B for more details about the parameter values used in Expokit), which we consider to be the "true" solution.[2] On the other hand, using Eq. (3.2.4) with $\tau = 0.01$ days, the IE method took a mere 53 seconds of CPU time, achieving a smaller (by a factor of 2.8) final L2 error of 5.35×10^{-4}. We can achieve a further reduction of the L2 error by using the RIE method with fixed step-size. This is clear from the results summarized in Table 3.1. We can achieve this performance however at the expense of increasing the CPU time required to calculate the solution. Note that we may be able to decrease the CPU time by using the RIE method with variable step-size (see Table 3.1). This method however results in a noticeable decrease of accuracy (at least for the example considered here), with an L2 error that is 2.8 times larger than the one obtained with the KSA method.

Since $R(t) = Q - S(t) - I(t)$, it suffices to focus on the joint probability mass function $\Pr[S(t), I(t)]$ of susceptible and infected pupils. It turns out however that the epidemic-free state occurs with high probability $\Pr[S(t), I(t) = 0]$, a situation that visually obscures the values of

[2] To be compatible with Expokit, we report here the L2 error. Note however that the error analysis of our method, provided in Appendix B, is based on the L1 error.

Table 3.1: The L2 error and CPU time associated with the four numerical solution methods of the master equation associated with the SIR model.

Numerical Method	**L2 Error**	**CPU Time**
KSA	1.48×10^{-3}	4328 seconds
IE	5.35×10^{-4}	52 seconds
RIE (fixed step-size)	1.11×10^{-4}	189 seconds
RIE (variable step-size)	4.06×10^{-3}	124 seconds

$\Pr[S(t), I(t)]$. For this reason, instead of $\Pr[S(t), I(t)]$, we depict in Fig. 3.2 a snapshot of the calculated joint conditional probability mass function $\Pr[S(t), I(t) \mid I(t) > 0]$ of the susceptible and infected pupils at the end of the 6th day, given that at least one pupil is infected. We have obtained this and all subsequent results by exclusively using the basic IE method.

In Fig. 3.3, we depict the dynamic profiles of the mean numbers of susceptible, infected and recovered pupils (solid green lines) as well as the the dynamic profiles of the ± 1 standard deviations (dashed red lines), computed directly from the joint probability mass function $\Pr[S(t), I(t), R(t)]$. We also depict the observed data (blue circles) obtained from the literature [135]. These results are identical to the results obtained by Monte Carlo estimation based on 1,000 trajectories sampled from the master equation using the Gillespie algorithm (only data related to the infected pupils are shown), and assures that the IE method produces the correct results. Unfortunately, we cannot employ the Gillespie algorithm to accurately estimate the joint probability mass function $\Pr[S(t), I(t), R(t)]$ in a reasonable time, due to the prohibitively large number of samples required by this method.

The bimodal nature of the probability mass function depicted in Fig. 3.2 clearly demonstrates that the LNA method used previously [17] is not appropriate for this model, since the method leads to a unimodal Gaussian approximation. As a matter of fact, the exact results depicted in Fig. 3.3 are different than the mean and standard deviation profiles depicted in Figures 3-4

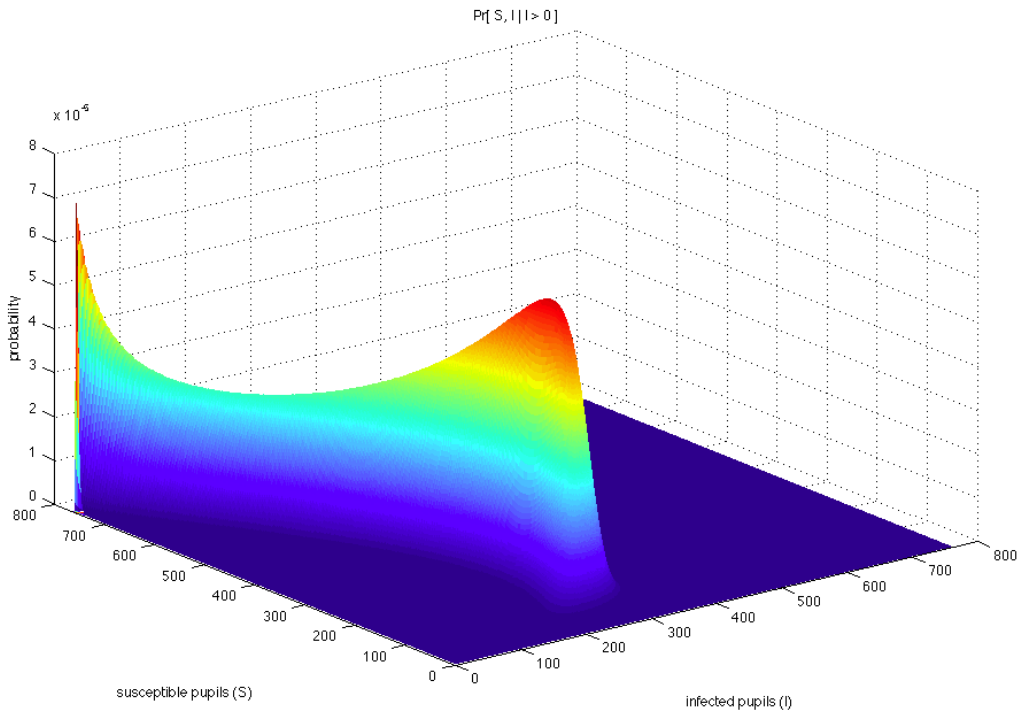

Figure 3.2: Joint conditional probability mass function $\Pr[S(t), I(t)|I(t) > 0]$ of susceptible and infected pupils at the end of the 6th day of the influenza epidemic.

in [17]. Because of the Gaussian nature of the LNA method, the previously reported results [17] over-estimate the means and under-estimate the standard deviations, since this technique is blind to the bimodal nature of the probability distribution. As a matter of fact, using the means and standard deviations to characterize the stochastic properties of individual classes in the SIR model is not appropriate. This is also evident by the fact that the ± 1 standard deviations can take negative values as well as values greater than 763. In Fig. 3.3, we have truncated these misleading values.

We can use the calculated joint probability mass functions $\Pr[S(t), I(t), R(t)]$ to study a number of dynamic properties of the SIR model in a stochastic setting. In Fig. 3.4(a), for example, we depict the evolution of the expected number of recovered pupils (solid green line), as well as the ± 1 standard deviations (dashed red lines), given that at least one pupil is always infected. During the first few days, few infections occur, and the expected number of recovered pupils will almost be

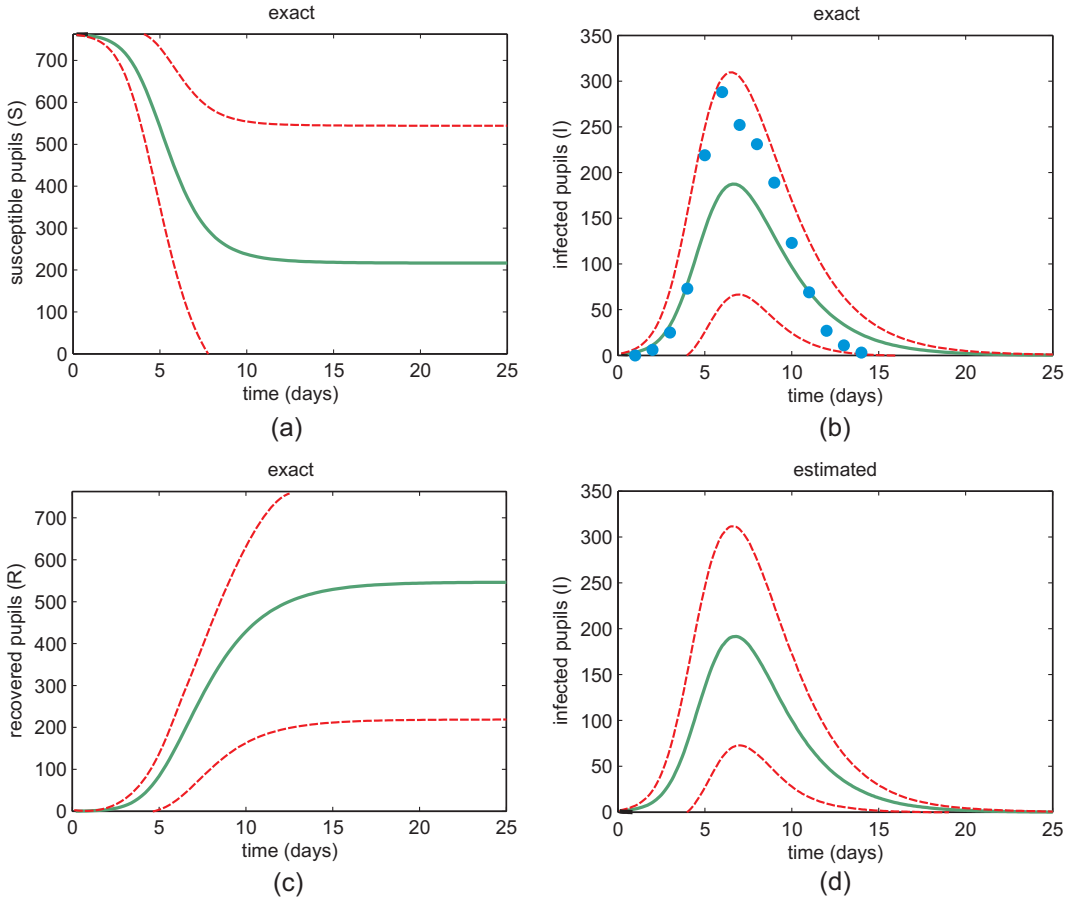

Figure 3.3: Dynamics of the mean profiles (solid green lines) and the ±1 standard deviation profiles (dashed red lines) of: (a) susceptible, (b) infected, and (c) recovered pupils. Monte Carlo estimates of the mean and standard deviation profiles of the infected pupils are depicted in (d). Blue circles in (b) mark available data.

zero. Subsequently, this number increases monotonically to 763, following a near sigmoidal curve. The ±1 standard deviation curves and the evolution of the Fano factor (variance/mean) depicted in Fig. 3.4(b), indicate that there is appreciable fluctuation in the number of recovered pupils during days 3–10, after which most pupils recover from the infection. According to the results depicted in Fig. 3.4(b), the maximum fluctuation in the number of recovered pupils occurs during the 6th day.

In Fig. 3.4(c), we depict the dynamic evolution of the calculated probability of extinction $\Pr[I(t) = 0]$, $t > 0$, during a period of 50 days. This evolution is characterized by four phases. During phase I (days 1–4), the probability of extinction increases rapidly from 0% to about 26%, due to the

small number of infectious pupils. During phase II (days 5–17), the probability remains relatively constant to about 26%. During this period of time, the epidemic takes its natural course, increasingly infecting susceptible individuals, who eventually recover from the disease. As a consequence, we do not expect the probability of extinction to increase during this phase. On the other hand, during phase III (days 18–40), the number of infected pupils monotonically decreases to zero. It is therefore expected that, during this phase, the probability of extinction will monotonically increase to its maximum value of one. Finally, during phase IV (days 40–50), there is no infectious pupils present. As a result, the influenza virus cannot be transmitted to the remaining susceptible pupils and the epidemic ceases to exist.

When studying an epidemic model with extinction, a task of practical interest is to calculate the number of individuals that escape infection. This is usually done by evaluating the expected number $\overline{e}$ of individuals that escape infection (or the average number of susceptible individuals that remain after extinction) as the mean value of the steady-state probability mass function $\Pr[S(\infty), I(\infty) = 0]$ [92]. The steady-state probability $\Pr[S(\infty), I(\infty) = 0]$ associated with our problem is depicted in Fig. 3.4(d). It turns out that $\overline{e} = 546.55$ in our case. Note however that, due to the bimodal nature of $\Pr[S(\infty), I(\infty) = 0]$, calculating $\overline{e}$ is misleading. However, by using the result depicted in Fig. 3.4(d), we can confirm that there is a 73.35% chance that 40 pupils or less, and a 26.53% chance that 753 pupils or more, escape infection. Clearly, these "confidence intervals" provide a more accurate statistical assessment of the number of individuals that escape infection than $\overline{e}$. Interestingly, there is only 0.12% chance that the number of pupils escaping infection is within the range $[41, 752]$, which includes the value of $\overline{e}$.

3.4 Discussion

Modeling the stochastic dynamics of a disease that spreads through a small and well-mixed population of individuals is an increasingly important subject of modern epidemiology. Unfortu-

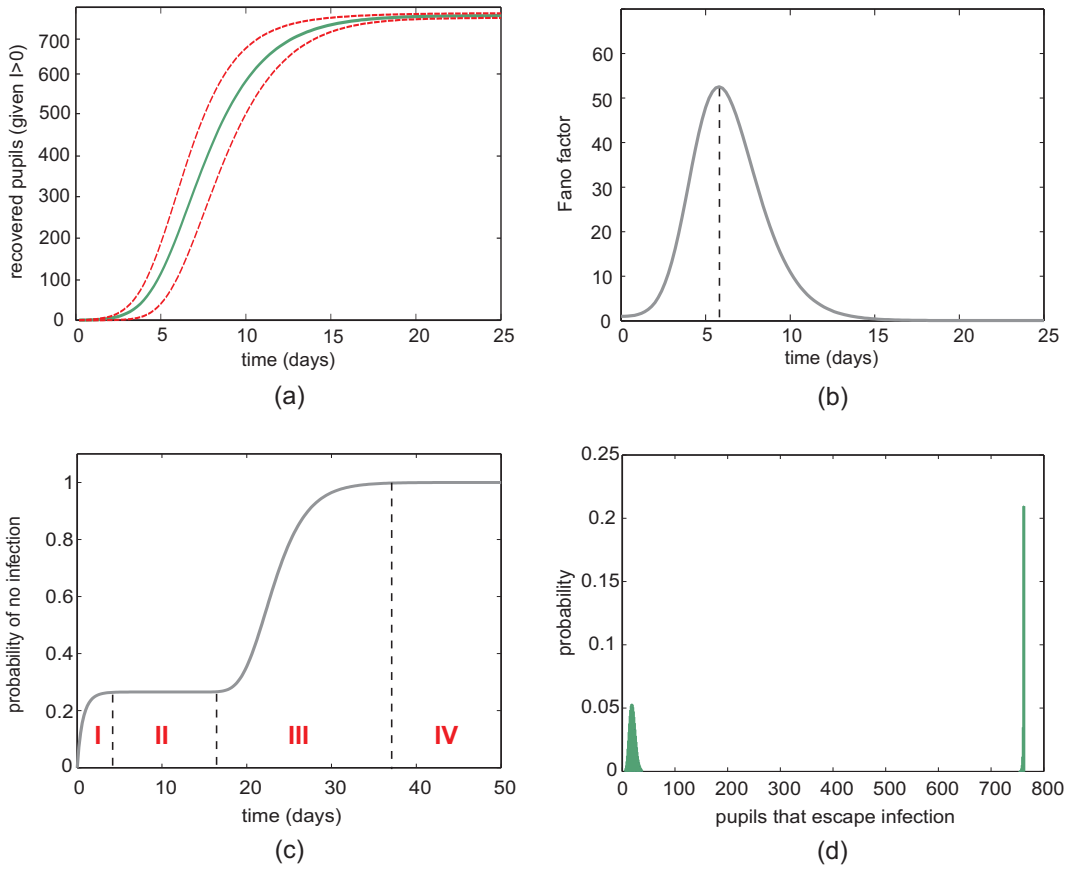

Figure 3.4: (a) Dynamic evolution of the expected number of recovered pupils (solid green line) and the ± 1 standard deviations (dashed red lines), given that at least one pupil is always infected. (b) The Fano factor (variance/mean) associated with the results in (a) as a function of time. (c) Dynamic evolution of the probability of extinction $\Pr[I(t) = 0]$, $t > 0$. (d) The approximation to the steady-state probability mass function $\Pr[S(\infty), I(\infty) = 0]$, given by the solution at 50 days.

nately, even for the simplest model, calculating the underlying probability distribution is a daunting task.

In an effort to address this problem, we have introduced in this chapter a new approach to numerically compute the probability mass function of a Markovian population process governed by the master equation. Implementation of this approach is feasible when the number of possible states is not prohibitively large. In this case, the proposed method can lead to *exact* statistical analysis of certain Markov models of interest, such as the SIR epidemic model.

The method introduced in this chapter is linear – both in terms of memory and compu-

tational requirements – with respect to the cardinality K of the sample space $\mathcal{Z}$ of the degrees of advancement of the underlying reactions. As a consequence, the method is feasible anytime $\mathcal{Z}$ is relatively small. In general, however, the cardinality of $\mathcal{Z}$ may grow arbitrarily large, making implementation of the method impossible without an appropriate FSP approximation [30]. Thus, the proposed technique is only applicable to models that constrain the number of reaction events, such as the SIR epidemic model considered in this chapter, or models for which the number of reaction events is sufficiently small during a time period of interest (i.e., models without "fast" reactions). Moreover, due to the well-known problem of the "curse of dimensionality," K grows exponentially with respect to the number of reactions M. Hence, models with many reactions cannot be solved by the proposed method.

An effort is currently underway to reduce the size of the sample space $\mathcal{Z}$, without compromising accuracy. A plausible way to accomplish this goal is to reduce the number of reactions involved by removing "fast" reactions using a multi-scale approximation technique, such as one of the techniques introduced for biochemical reaction systems [11,12,137], and to adaptively update $\mathcal{Z}$ at each time point t by confining it to the smallest possible subspace $\mathcal{Z}(t)$ of $\mathcal{Z}$. Because of the lower-triangular and sparse nature of matrix $\mathbb{Q}$ in Eq. (3.2.4), it is also plausible that we employ optimized algorithms developed for solving sparse triangular systems of linear equations on parallel and distributed memory computer architectures [138], indicating that future efforts towards solving the master equation could potentially focus on using high-performance computing systems.

Finally, we note that an earlier work [139] in the mathematics literature has shown how a general Markov process on a countable state-space may be mapped to another stochastic process on an augmented state-space such that the generator matrix of this augmented process is triangular. This result is more general than our use of the DA process, but for the reactive processes considered in this chapter that are governed by Eq. (2.1.1) the DA process is preferable, as discussed in Appendix B. Although this earlier work identified and exploited the triangularity of the new stochastic process, it used a recursive solution technique [140, 141] which turns out to be inferior to the IE method

developed in this chapter, as discussed in Appendix B. Intuitively, the IE method is superior because these earlier works exploit only the triangularity and not the sparseness of systems governed by Eq. (2.1.1). Future developments on the IE method should thus maintain a focus on both the triangularity and the sparsity of the underlying DA process.

Chapter 4

Statistical Testing of Master Equation Approximations

Due to the nonlinear nature of most reactions and the large size of the underlying state-spaces, computing the exact solution of the master equation is intractable in general. For this reason, a number of approximation techniques have been proposed in the literature to deal with this problem. Unfortunately, it is not easy to check whether a given approximation technique produces acceptable results. As a consequence, approximating the solution of the master equation may lead to significant errors without any prior warning. Being aware of this serious problem, we would like to investigate whether a particular technique produces an acceptable approximation to the solution of the master equation and act accordingly. If our prior investigation leads to the conclusion that the method is acceptable with some level of confidence, then we can proceed using it with possibly substantial computational savings. If not, then we may try to develop a more appropriate approximation method or commit substantial computational resources to obtain exact results if possible.

In this chapter[1], we propose a hypothesis testing framework that allows us to reject an

[1]Materials in this chapter are reprinted with permission from "Statistically testing the validity of analytical and computational approximations to the chemical master equation", by Garrett Jenkinson and John Goutsias, *Journal of Chemical Physics*, vol. 138, issue 20, number 204108, Copyright 2013, AIP Publishing LLC.

approximation technique when it is not valid, or determine that it is safe to use with a predefined level of confidence. By drawing a moderate number of samples from the master equation, the proposed methods employ the well-known Kolmogorov-Smirnov statistic to test the validity of a given approximation technique. If we cannot reject the approximation, then our methods provide a quantitative measure of the extent to which we can trust it. Our approach is general enough to deal with any master equation and can be used to test the validity of any analytical approximation method or any approximative sampling technique of interest.

4.1 Motivation

As we mentioned before, a popular analytical approximation to the solution of the master equation is obtained by the LNA method, which turns out to be accurate in systems of sufficiently large volume at small enough times [1]. Unfortunately, there are no quantitative guidelines for determining whether the LNA method produces a valid approximation to the solution of the master equation for a given system volume and time. The only available guidelines are for master equations with linear propensity functions, [142] for which the exact solution can be easily computed [28]. Despite this uncertainty, the LNA method has been extensively and sometimes incorrectly used to study the behavior of chemical reaction systems, as well as epidemiological, ecological, social, and neural networks [5].

When sufficiently accurate analytical approximations to the solution of the master equation are not possible, one may employ computational techniques to sample the master equation and use the resulting samples to compute Monte Carlo estimates of various moments and joint probability distributions. Although exact sampling of the master equation is possible by means of the Gillespie algorithm, [14,45,46] this method can be computationally very demanding, especially when estimating high-order statistical summaries or joint probability distributions. As a consequence, one must rely on techniques that draw approximative samples from the master equation in a computationally

efficient way. It turns out that there is no shortage of approximative sampling methods, [5] with each method having its own advantages and disadvantages as well as its own parameters which must be appropriately tuned to obtain acceptable estimation performance. The main problem here is not lack of approximative sampling methods, but lack of confidence in using these methods. Again, it is not in general possible to check whether a particular scheme produces legitimate samples and, currently, there is no effective methodology to address this problem.

We should note here that a method has been proposed in the literature for measuring the accuracy of approximative methods for sampling the master equation using the Kolmogorov-Smirnov distance between two cumulative probability distributions [143]. Although a brief mention to hypothesis testing is made in that publication (mainly to justify the use of the Kolmogorov-Smirnov distance), the authors fall short of developing a statistically rigorous method for checking the validity of a given approximation method. Instead, they focus their interest on using the Kolmogorov-Smirnov distance to study sensitivity properties of stochastic chemical reaction systems [144], as well as to measure convergence properties of Poisson leaping [145], a popular algorithm for drawing approximative samples from the master equation. On the other hand, the work presented in this chapter is based on using the Kolmogorov-Smirnov distance to develop rigorous hypothesis testing approaches for rejecting or accepting analytical approximation methods or approximative sampling techniques by making statistical decisions about the validity of a given scheme. More details related to the material presented in this chapter can be found in [146].

4.2 LNA for the population process

In many cases of interest, we can find a parameter Ω that measures the relative size of stochastic fluctuations, such that fluctuations become increasingly smaller as Ω becomes larger. For example, in chemical reaction systems, Ω is often taken to be the system volume. In such systems, stochastic fluctuations gradually diminish as the system approaches the *thermodynamic limit*

(denoted by $\Omega \to \infty$) for which Ω increases to infinity in a manner that keeps the concentrations of the underlying chemical species fixed. In this case, it is intuitive to expect that the probability of a reaction to occur within an infinitesimally small time interval $[t, t + dt)$ depends on the "concentration"

$$\widetilde{\boldsymbol{X}}(t) := \frac{\boldsymbol{X}(t)}{\Omega} \tag{4.2.1}$$

of the population process at time t and that this probability does not change when Ω varies as long as the concentrations remain fixed [1]. This implies that the propensity functions $\pi_m(\boldsymbol{x})$ must only depend on the concentrations $\boldsymbol{x}/\Omega$. It is commonly assumed that

$$\pi_m(\boldsymbol{x}) = f(\Omega)\big[\widetilde{\pi}_m(\boldsymbol{x}/\Omega) + \Omega^{-1}\widetilde{\pi}'_m(\boldsymbol{x}/\Omega) + \cdots\big], \tag{4.2.2}$$

for some nonnegative functions f, $\widetilde{\pi}_m$, $\widetilde{\pi}'_m, \ldots$ which do not depend on Ω [1].

Under the previous assumptions, the law of large numbers implies that $\widetilde{\boldsymbol{X}}(t)$ converges, as $\Omega \to \infty$, to the solution $\boldsymbol{\mu}(t)$ of the standard reaction rate (macroscopic) equations [74]

$$\frac{d\mu_n(t)}{dt} = \sum_{m \in \mathcal{M}} s_{nm}\widetilde{\pi}_m(\boldsymbol{\mu}(t)), \quad \text{for } n \in \mathcal{N}, \tag{4.2.3}$$

initialized by $\boldsymbol{\mu}(0) = \boldsymbol{x}_0/\Omega$, with $\mu_n(t)$ being the n-th element of vector $\boldsymbol{\mu}(t)$. Moreover, and for sufficiently large Ω, the concentration process can be approximated by using the following *ansatz*:

$$\widetilde{X}_n(t) = \mu_n(t) + \frac{1}{\sqrt{\Omega}}\,\Xi_n(t), \quad \text{for } \; t > 0, \;\; n \in \mathcal{N}, \tag{4.2.4}$$

where $\Xi_n(t)$ is a noise component that quantifies the fluctuations associated with the molecular concentrations. For each Ω, Eq. (4.2.4) decomposes the random variable $\widetilde{X}_n(t)$ into a *macroscopic* (deterministic) component $\mu_n(t)$ and an additive noise component $\Xi_n(t)$ that is independent of Ω. This ansatz is based on the premise that fluctuations decrease at a rate proportional to $\Omega^{-1/2}$. Many justifications have been offered in the literature to explain the accuracy of this assumption for systems close to the thermodynamic limit [1, 74, 75, 147].

It can be shown that, for sufficiently large Ω, the dynamic evolution of the probability density function of the $N \times 1$ noise vector $\boldsymbol{\Xi}(t)$ is approximately governed by a *linear* Fokker-Planck

equation [1]. By solving this equation, one finds that $\boldsymbol{\Xi}(t)$ is a zero mean Gaussian process, whose covariances $c_{nn'}(t) := \mathrm{E}[\Xi_n(t)\Xi'_n(t)]$ solve the following Lyapunov equations:

$$\frac{dc_{nn'}(t)}{dt} = \sum_{k\in\mathcal{N}}\Big(\sum_{m\in\mathcal{M}} s_{nm}g_{mk}(t)\Big)c_{kn'}(t) + \sum_{k\in\mathcal{N}}\Big(\sum_{m\in\mathcal{M}} s_{n'm}g_{mk}(t)\Big)c_{kn}(t) + \sum_{m\in\mathcal{M}} s_{nm}s_{n'm}\widetilde{\pi}_m(\boldsymbol{\mu}(t)), \tag{4.2.5}$$

for $t > 0$, $n, n' \in \mathcal{N}$, initialized by $c_{nn'}(0) = 0$, where $g_{nn'}(t) := [\partial\widetilde{\pi}_n(\boldsymbol{y})/\partial y_{n'}]_{\boldsymbol{y}=\boldsymbol{\mu}(t)}$ are the derivatives of the propensity functions $\widetilde{\pi}$ evaluated at the solution of the macroscopic equations (4.2.3). These covariances can be stored into a covariance matrix $\mathbb{C}(t)$.

Finally, by multiplying both sides of Eq. (4.2.4) with Ω, one finds that the solution to the master equation $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ can be approximated by the multivariate Gaussian distribution

$$\widehat{P}(\boldsymbol{x};t) = \frac{1}{\sqrt{(2\pi)^N\Omega|\mathbb{C}(t)|}}\exp\left\{-\frac{1}{2\Omega}\big[\boldsymbol{x}-\Omega\boldsymbol{\mu}(t)\big]^T\mathbb{C}^{-1}(t)\big[\boldsymbol{x}-\Omega\boldsymbol{\mu}(t)\big]\right\}, \tag{4.2.6}$$

with mean vector $\Omega\boldsymbol{\mu}(t)$ and covariance matrix $\Omega\mathbb{C}(t)$, where $\boldsymbol{\mu}(t)$ solves the macroscopic equation (4.2.3) and $\mathbb{C}(t)$ solves the Lyapunov equation (4.2.5). In Eq. (4.2.6), $|\mathbb{A}|$ denotes the determinant of matrix $\mathbb{A}$.

Note that computation of the LNA $\widehat{P}(\boldsymbol{x};t)$ requires that we first solve the N macroscopic equations (4.2.3) and then solve the $N(N+1)/2$ Lyapunov equations (4.2.5). The main drawback of this approximation however is the absence of a technique that allows us to determine, with some level of statistical confidence, whether $\widehat{P}(\boldsymbol{x};t)$ is indeed an acceptable approximation to the true solution $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ of the master equation for a given system size Ω. We propose such a method next.

4.3 Testing the validity of analytical approximations

Given an analytical approximation $\widehat{P}(\boldsymbol{x};t)$ to the solution $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ of the master equation (such as the one obtained by the LNA method), we would like to investigate, in a statistically rigorous manner, whether this approximation is valid in some useful sense. In this section, we propose a simple procedure to address this problem based on hypothesis testing.

The main idea behind our method is to sample the master equation using *exact* sampling in order to produce a moderate amount of population data drawn from the (unknown) probability distribution $p_{\boldsymbol{X}}(\boldsymbol{x};t)$ that solves the master equation. We subsequently use this data, in an appropriately designed statistical test, to reject the approximation, if the test decides that the data have not been sampled from $\widehat{P}(\boldsymbol{x};t)$, or accept $\widehat{P}(\boldsymbol{x};t)$ as being a valid approximation to the solution of the master equation in some sense, if the test fails to reject the possibility that the data have not been drawn from $\widehat{P}(\boldsymbol{x};t)$. In the latter case, we quantify our trust in $\widehat{P}(\boldsymbol{x};t)$ being a good approximation to the solution of the master equation by using appropriately constructed confidence regions. To simplify our discussion, we begin by focusing on the case when the master equation is one-dimensional. We then extend our method to the case of multiple dimensions.

4.3.1 The one-dimensional case

Hypothesis testing

Let us denote by $\{x_1(t), x_2(t), \ldots, x_L(t)\}$ statistically independent samples of the one-dimensional population process $X(t)$ drawn at time t from the master equation using the *exact* Gillespie algorithm. Based on these samples, we can calculate the *empirical* cumulative distribution function (CDF) $G(x;t)$, given by

$$G(x;t) = \frac{1}{L}\sum_{l=1}^{L}[x_l(t) \leq x], \tag{4.3.7}$$

where $[\,]$ is the Iverson bracket.[2] This CDF provides an approximation to the CDF $F(x;t)$ that corresponds to the true but unknown solution $p_{\boldsymbol{X}}(x;t)$ of the master equation. On the other hand, let $\widehat{F}(x;t)$ be the CDF that corresponds to the analytical approximation $\widehat{P}(x;t)$ of the solution of the master equation. To measure how good the analytical approximation is, we may calculate the following "distance"

$$S(t) = \max_{x}\left|G(x;t) - \widehat{F}(x;t)\right| \tag{4.3.8}$$

[2] $[a \leq b]$ equals 1, if $a \leq b$, and 0 otherwise.

between the two CDFs, known as the *Kolmogorov-Smirnov* (KS) *statistic* [148, 149]. This statistic computes the largest absolute difference between the CDF $\widehat{F}(x;t)$ produced by the analytical approximation and the empirical CDF $G(x;t)$ observed when sampling the master equation using Gillespie's *exact* method. It is known that, as the number of samples L increases to infinity, the empirical CDF $G(x;t)$ will converge almost surely to the true CDF $F(x;t)$, for every x. Therefore, and for sufficiently large L, the KS statistic can provide an effective way of evaluating the accuracy of the analytical approximation. In particular, "small" values of $S(t)$ below a threshold s_0 may indicate that there is little difference observed between the approximating analytical and empirical distributions, in which case, we may claim that the analytical approximation method provides an acceptable approximation to the solution of the master equation. On the other hand, values of $S(t)$ above s_0 may indicate that there is significant discrepancy between the observed samples and the prediction of the analytical approximation method, in which case, we may reject the probability distribution obtained by the analytical approximation.

To implement the previous approach in a statistically rigorous manner, we must formulate it as a hypothesis testing problem that deals with the following *null* and *alternative* hypotheses:

H_0: $P = \widehat{P}$

H_A: $P \neq \widehat{P}$,

where we use the shorthand P for the solution $p_{\boldsymbol{x}}(x;t)$ of the master equation at time t and denote by $\widehat{P}$ the corresponding probability distribution $\widehat{P}(x;t)$ suggested by the analytical approximation method at hand. This means that the set $\mathcal{P}_A = \{Q : Q \neq \widehat{P}\}$ of all probability distributions that satisfy the alternative hypothesis contains an infinite number of elements, whereas, the corresponding set $\mathcal{P}_0 = \{\widehat{P}\}$ for the null hypothesis has only one element. We refer to a probability distribution Q in $\mathcal{P}_A$ as an *alternative* distribution and call $\widehat{P}$ the *null* distribution. Note that the data samples $\{x_l(t),\ l = 1, 2, \ldots, L\}$ used to determine the empirical CDF $G(x;t)$, given by Eq. (4.3.7), are drawn *implicitly* from the *true* distribution P.

We can use the KS statistic to implement the previous hypothesis testing problem, with the null hypothesis assuming that the data $\{x_l(t),\ l = 1, 2, \ldots, L\}$ are drawn from the null distribution $\widehat{P}(x;t)$ specified by the analytical approximation method. If we could reject this hypothesis with a small p-value then, we could conclude that the analytical approximation method does not produce an acceptable approximation to the solution of the master equation at time t. The KS statistic has been extensively studied in the literature and most statistical software packages provide the p-value for a given value of S. If the resulting p-value is less than a chosen significance level α, then we reject the null hypothesis; i.e., we reject the validity of the analytical approximation method. Otherwise, we conclude that the data are not sufficiently persuasive to lead us to the conclusion that the analytical approximation is not valid.

It turns out that the null hypothesis is rejected when the KS statistic $S(t)$ is larger than a critical value $s_0(\alpha)$ that depends on the chosen significance level α. This implies that $\mathcal{D}_\alpha(t) := \{\mathcal{D}_\alpha(x;t), -\infty < x < \infty\}$ is the *decision band* of the test, where

$$\mathcal{D}_\alpha(x;t) = \big(\max\{0, \widehat{F}(x;t) - s_0(\alpha)\}, \min\{1, \widehat{F}(x;t) + s_0(\alpha)\}\big), \tag{4.3.9}$$

so the test will reject the analytical approximation when values of the empirical distribution $G(x;t)$ lie outside of $\mathcal{D}_\alpha(t)$ for at least one x. However, when all values of $G(x;t)$ fall entirely within $\mathcal{D}_\alpha(t)$, then we may claim with $(1-\alpha)\%$ confidence that the CDF obtained by analytical approximation is close to the true CDF, in a sense quantified by the "width"[3] $2s_0(\alpha)$ of the *confidence band* $\mathcal{C}_\alpha(t) := \{\mathcal{C}_\alpha(x;t), -\infty < x < \infty\}$, where $\mathcal{C}_\alpha(x;t)$ is the $(1-\alpha)\%$ confidence interval for the true CDF value $F(x;t)$,[4] given by (this is a direct consequence of Example 4.4.6 in Ref. [149])

$$\mathcal{C}_\alpha(x;t) = \big(\max\{0, G(x;t) - s_0(\alpha)\}, \min\{1, G(x;t) + s_0(\alpha)\}\big). \tag{4.3.10}$$

We say in this case that we are at least $(1-\alpha)\%$ certain that the analytical approximation provides an acceptable solution to the master equation. It can be shown that, when $\widehat{F}(x;t)$ is a continuous

[3]Note that the actual widths of the decision and confidence bands are both less than or equal to $2s_0(\alpha)$.

[4]This means that, with probability at least $1-\alpha$, one may claim that the true (but unknown) CDF value $F(x;t)$ of the solution to the master equation at population x and time t is contained within the confidence interval $\mathcal{C}_\alpha(x;t)$.

function of x (which is the case when considering the LNA method), the critical value $s_0(\alpha)$ can be well approximated by [150]

$$s_0(\alpha) = \frac{\sqrt{|\ln(\alpha/2)/2|}}{\sqrt{L} + 0.12 + 0.11/\sqrt{L}} \,. \tag{4.3.11}$$

When $\widehat{F}(x;t)$ is not a continuous function of x (e.g., when the analytical approximation to the master equation is a discrete distribution), the critical value of the KS test is always smaller than the critical value given by Eq. (4.3.11) [148, 151, 152]. In this case, hypothesis testing will be conservative, in the sense that the probability of the Type I error (i.e., the error incurred by rejecting the validity of the analytical approximation method when the solution of the master equation is given by the probability distribution derived by this method) is guaranteed to be less than or equal to the significance level α, but the test may be characterized by a larger probability of Type II error (i.e., the error of failing to reject the analytical approximation when this approximation does not solve the master equation) and thus have lower power.[5] However, if this turns out to be a problem, one can estimate the true critical value by Monte Carlo simulation using the simple procedure described in Appendix C.

We should point out here that we cannot use the previous hypothesis testing procedure to plainly accept the null hypothesis (i.e., to claim that the approximation method provides a valid solution to the master equation). In fact, hypothesis testing can only lead to rejecting the null hypothesis or to failing to reject the null hypothesis. On the other hand, computing a confidence band allows us to accept the validity of an approximation to the solution of the master equation with $(1-\alpha)\%$ confidence. Although a confidence band cannot be used to justify acceptance of the null hypothesis, it can provide strong evidence that an approximation technique is sufficiently accurate for all practical purposes. This is a direct consequence of the fact that the true CDF is necessarily within the confidence band. Therefore, if the width of the confidence band is sufficiently small, then we can claim with $(1-\alpha)\%$ confidence that the approximating CDF is close to the true CDF and accept the approximation with that level of confidence.

[5] The power of a statistical test equals 1 minus the probability of the Type II error.

We should finally note here that [see Eq. (4.1.3) in Ref. [149]]

$$S(t) = \max_{l=1,\ldots,L} \left\{ \max\left\{ \frac{l}{L} - \widehat{F}(x_{[l]}(t);t), \widehat{F}(x_{[l]}(t);t) - \frac{l-1}{L} \right\} \right\}, \tag{4.3.12}$$

where $x_{[1]}(t) \leq x_{[2]}(t) \leq \cdots \leq x_{[L]}(t)$ is the ordered observed sample. Moreover,

$$\widehat{F}(x;t) = \Phi\left(\frac{x - \Omega\mu_1(t)}{\sqrt{\Omega c_{11}(t)}} \right) \tag{4.3.13}$$

for the LNA method, where Φ is the CDF of the standard normal distribution and $\mu_1(t)$ is found by solving the one-dimensional version of Eq. (4.2.3) while $c_{11}(t)$ is found according to the one-dimensional version of Eq. (4.2.5). As a consequence, the KS statistic for the LNA method is given by

$$S(t) = \max_{l=1,\ldots,L} \left\{ \max\left\{ \frac{l}{L} - \Phi\Big(\frac{x_{[l]}(t) - \Omega\mu_1(t)}{\sqrt{\Omega c_{11}(t)}} \Big), \Phi\Big(\frac{x_{[l]}(t) - \Omega\mu_1(t)}{\sqrt{\Omega c_{11}(t)}} \Big) - \frac{l-1}{L} \right\} \right\}. \tag{4.3.14}$$

Equation (4.3.12) and Eq. (4.3.14) provide formulas for the efficient implementation of the KS test statistic in the general and LNA cases, respectively.

Choosing the significance level and sample size

The previous KS hypothesis testing approach requires two parameters to be specified: the significance level α and the sample size L. These parameters affect the performance of the test, both in terms of error rates and computational efficiency, and their values must be chosen carefully. In the following, we discuss possible strategies for determining appropriate values for these parameters in three different scenarios of interest.

Scenario 1: When sufficient resources are available to perform exact sampling, we must focus on choosing values for α and L so that the hypothesis test meets specified performance criteria. For example, we may limit the probability of the Type I error to some allowable value and set α equal to that value, since the probability of the Type I error is never larger than α. Moreover, we may constrain the width $2s_0(\alpha)$ of the confidence band $\mathcal{C}_\alpha(t)$ to be at most as large as a desirable size

w_0, in which case we approximately obtain, by virtue of Eq. (4.3.11), that

$$L \geq \left\lceil \frac{2|\ln(\alpha/2)|}{w_0^2} \right\rceil, \tag{4.3.15}$$

where $\lceil x \rceil$ is the ceiling function. To improve computational efficiency, we may accept the lower bound and set $L = \lceil 2|\ln(\alpha/2)|/w_0^2 \rceil$.

In this case, if the KS test fails to reject the analytical approximation, we can conclude with $(1-\alpha)\%$ confidence that, for every x, the CDF $\widehat{F}(x;t)$ suggested by the analytical approximation is within a band of width w_0 from the true CDF $F(x;t)$. As a consequence, we may conclude that the analytical approximation method provides an acceptable approximation to the solution of the master equation.

Scenario 2: In large or stiff chemical reaction networks, sampling the master equation using the exact Gillespie algorithm can be computationally demanding. In this case, the value of L will be limited, by available computational resources, to a rather small maximum acceptable value. On the other hand, to control the probability of the Type I error, we may specify a maximum allowable value for this probability as before and set the significance level α equal to that value.

If L is small and the KS test fails to reject the analytical approximation, the decision band $\mathcal{D}_\alpha(t)$ may be unacceptably wide [note that $s_0(\alpha) \propto L^{-1/2}$]. This indicates that we must draw more data samples from the master equation to improve testing performance and our confidence about the validity of the analytical approximation. In this case, and without additional samples, nothing more can be said about the accuracy of the method, and we conclude that the analytical approximation must be used with caution.

Scenario 3: In the previous two scenarios, the value of the significance level is chosen so that the probability of the Type I error is small. On the other hand, we would also like the probability of the Type II error to be small as well. This is equivalent to requiring that the power of the test against the alternative distribution is large. Therefore, we may attempt to determine values for α and L so that the KS hypothesis testing procedure results in a large power over all alternative

probability distributions in $\mathcal{P}_A$. This however is not possible, since we cannot specify every single probability distribution in $\mathcal{P}_A$. Instead, we may address the previous problem over a parametric subclass $\mathcal{Q}_A = \{Q_r, 0 \leq r \leq 1\} \subset \mathcal{P}_A$ of known alternative probability distributions, such that $Q_r \simeq \widehat{P}$, for large r, and $Q_r \not\simeq \widehat{P}$, for small r. In this case, we would like the power of the KS test against an alternative distribution Q_r to be sufficiently small for large values of r and sufficiently large for small values of r.

An example that illustrates this case is when the macroscopic system described by Eq. (4.2.3) is bistable, indicating that, for sufficiently large system sizes Ω, the solution of the master equation will be bimodal at steady-state, with the two modes located at $\Omega\mu^*$ and $\Omega\mu^{**}$, where μ^*, μ^{**} are the stable fixed points of the macroscopic equation [153]. If the system is initialized within the basin of attraction of one of these stable fixed points, say $\Omega\mu^*$, then the resulting dynamics may be well characterized, during at least some initial period of time, by a Gaussian distribution that moves towards $\Omega\mu^*$ as time progresses. For sufficiently large Ω, this distribution will place appreciable probability mass in the vicinity of $\Omega\mu^*$ and negligible probability mass around $\Omega\mu^{**}$, since the probability of moving towards $\Omega\mu^{**}$ exponentially diminishes to zero as the system approaches the thermodynamic limit [153]. We may therefore assume that the solution of the master equation is approximately given by

$$Q_r(x;t) = r\widehat{P}^*(x;t) + (1-r)\widehat{P}^{**}(x;t), \quad 0 \leq r \leq 1, \tag{4.3.16}$$

where r is the probability of initializing the master equation within the basin of attraction of $\Omega\mu^*$, $1-r$ is the probability of initializing the master equation within the basin of attraction of $\Omega\mu^{**}$, and $\widehat{P}^*(x;t)$, $\widehat{P}^{**}(x;t)$ are the LNA distributions obtained by initializing the master equation within the basins of attraction of $\Omega\mu^*$ and $\Omega\mu^{**}$, respectively. It is now desirable to determine appropriate values for α and L so that the KS test almost always fails to reject the LNA approximation $\widehat{P}^*(x;t)$ when the solution of the master equation is approximately given by $Q_r(x;t) = \widehat{P}^*(x;t)$, whereas, it almost always rejects the LNA approximation $\widehat{P}^*(x;t)$ when the solution of the master equation is

approximately given by $Q_r(x;t) = \widehat{P}^{**}(x;t)$. We propose a method to do so in Appendix C.

4.3.2 Extension to multiple dimensions

Extending the previous method to the case of multiple dimensions is conceptually straightforward but mathematically more demanding. Unfortunately, the KS test does not directly generalize to random variables in more than one dimension, and there is no statistical test in a multivariate setting that stands out [154]. For the multivariate Gaussian setting, which is implied by the LNA method, a number of tests have been proposed in the literature [155], including extensions of the KS test to more than one dimension [156–158].

Here, we focus on the simplest multivariate extension of KS testing wherein the univariate test considered above is independently applied on each of the N one-dimensional marginal distributions. We justify this route by considering the fact that, in practice, we are most often interested in one-dimensional marginal distributions and statistical summaries of individual species populations. In addition, an approximation to the solution of the master equation, such as the one provided by the LNA method, is usually justified based on some theoretical argument (e.g., that the LNA method is valid for large systems). In such cases, if hypothesis testing demonstrates that the approximation produces marginal distributions that are close to the true distributions, then this combination of theoretical and empirical evidence allows us to accept the validity of the approximation method, even though hypothesis testing alone does not support such a decision.

An important advantage of the approach discussed in this subsection is the existence of confidence bands around the marginal distributions, which may not be available when other hypothesis tests are used. Decision and confidence regions around each marginal distribution can be constructed as in the univariate case, but care must be taken when combining the results from each marginal distribution into a single decision about the multidimensional goodness-of-fit of the analytical approximation method due to concerns associated with multiple-testing. We have decided to handle this issue by combining results using Tippett's method [159]. This method incorporates the

p-values from each of the N independent tests into a single p-value and leads to a global hypothesis testing procedure that deals with the following null and alternative hypotheses:

H_0: $P_n = \widehat{P}_n$, for every $n \in \mathcal{N}$

H_A: $P_n \neq \widehat{P}_n$, for at least one $n \in \mathcal{N}$.

Here, we use the shorthand P_n for the marginal distribution $P_n(x_n;t)$ of the multivariate solution $p_{\boldsymbol{x}}(\boldsymbol{x};t)$ of the master equation at time t and denote by $\widehat{P}_n$ the corresponding marginal probability distribution $\widehat{P}_n(x_n;t)$ implied by the analytical approximation method. This approach allows us to avoid complex issues associated with multiple testing. We provide details next.

For each $n = 1, 2, \ldots, N$, we first use the exact Gillespie algorithm to draw L independent vector samples $\{\boldsymbol{x}_l(t),\ l = 1, 2, \ldots, L\}$ from the master equation (2.1.8). We then use these samples to calculate the marginal empirical CDF

$$G_n(x;t) = \frac{1}{L}\sum_{l=1}^{L} [x_{l,n}(t) \leq x], \tag{4.3.17}$$

where $x_{l,n}(t)$ is the n-th element of vector $\boldsymbol{x}_l(t)$. Subsequently, we compute the KS statistic $S_n(t) = \max_x |G_n(x;t) - \widehat{F}_n(x;t)|$ associated with the n-th marginal distribution of the population process $\boldsymbol{X}(t)$ at time t. For the LNA method, this statistic is given by

$$S_n(t) = \max_{l=1,\ldots,L} \left\{ \max\left\{ \frac{l}{L} - \Phi\Big(\frac{x_{[l],n} - \Omega\mu_n(t)}{\sqrt{\Omega c_{nn}(t)}}\Big), \Phi\Big(\frac{x_{[l],n} - \Omega\mu_n(t)}{\sqrt{\Omega c_{nn}(t)}}\Big) - \frac{l-1}{L} \right\} \right\}, \tag{4.3.18}$$

where $x_{[1],n} \leq x_{[2],n} \leq \cdots \leq x_{[L],n}$ is the ordered observed sample in the n-th dimension. Note that the μ_n's solve Eq. (4.2.3), whereas, the c_{nn}'s are found by solving Eq. (4.2.5). We finally use the KS hypothesis testing procedure to compute the p-value $p_n(t)$ of the observed value of $S_n(t)$.

After computing all p-values $\{p_n(t), n \in \mathcal{N}\}$, we combine them into one p-value $p(t)$ using Tippett's method, in which case,

$$p(t) = 1 - \big(1 - \min_{n\in\mathcal{N}}\{p_n(t)\}\big)^N. \tag{4.3.19}$$

We then reject the multivariate analytical approximation at time t with significance level α whenever $p(t) < \alpha$ or (equivalently) whenever

$$\min_{n\in\mathcal{N}}\{p_n(t)\} < \alpha_0 := 1 - \sqrt[N]{1-a}. \tag{4.3.20}$$

If desired, we can calculate the decision and confidence intervals

$$\begin{aligned}\mathcal{D}^n_{\alpha_0}(x;t) &= \big(\max\{0, \widehat{F}_n(x;t) - s_0(\alpha_0)\}, \min\{1, \widehat{F}_n(x;t) + s_0(\alpha_0)\}\big)\\ \mathcal{C}^n_{\alpha}(x;t) &= \big(\max\{0, G_n(x;t) - s_0(\alpha)\}, \min\{1, G_n(x;t) + s_0(\alpha)\}\big),\end{aligned} \tag{4.3.21}$$

for $n = 1, 2, \ldots, N$, where $\widehat{F}_n(x;t)$ is the n-th marginal CDF of the analytical approximation and s_0 is given by Eq. (4.3.11).[6] In this case, the test will reject the analytical approximation when, for at least one n, values of the empirical distribution $G_n(x;t)$ lie outside of the decision band $\mathcal{D}^n_{\alpha_0}(t)$. However, when this is not true, we may claim with $(1-\alpha)\%$ confidence that, for each $n = 1, 2, \ldots, N$, the marginal CDF $\widehat{F}_n(x;t)$ obtained by the analytical approximation is close to the true marginal CDF $F_n(x;t)$, in a sense quantified by the width $2s_0(\alpha)$ of the confidence band $\mathcal{C}^n_{\alpha}(t)$.

4.4 Testing the validity of approximative sampling

Before we use an approximative sampling technique for a particular Monte Carlo estimation task, we should statistically test whether the technique provides a valid approximation to exact sampling. By drawing a number of exact and approximate samples from the master equation, we can gain statistical confidence for the accuracy of approximative sampling in a given setting. If the resulting confidence is high, we can utilize the sampling method in computationally more demanding tasks. Moreover, we can safely pool together the exact and approximate samples and reuse them in the Monte Carlo estimation problem at hand.

We address this problem by employing a *two-sample* version of the previously discussed KS hypothesis testing procedure, which we refer to as TSKS test [160, 161]. This is despite the

[6] In the multidimensional case, the decision band is computed using the significance level α_0, given by Eq. (4.3.20), and not the significance level α.

fact that more powerful multivariate statistical tests exist to deal with this task [162–164]. The main reason for our choice is the fact that the TSKS test provides confidence intervals around the marginal CDFs, whereas the more complex multivariate tests lack this capability. Moreover, the TSKS test is simpler to implement (it is a standard tool in most statistical software) and easier to understand, since it is a natural extension of the previously discussed KS test. As a consequence, we focus here on the case when only one-dimensional marginal distributions and statistical summaries of individual species populations are of interest. As before, we first discuss the one-dimensional case and then extend our discussion to multiple dimensions.

4.4.1 The one-dimensional case

Hypothesis testing

Let us denote by $\{x_1(t), x_2(t), \ldots, x_L(t)\}$ statistically independent samples drawn from the master equation at time t using exact sampling, whereas, let $\{\widetilde{x}_1(t), \widetilde{x}_2(t), \ldots, \widetilde{x}_L(t)\}$ be another set of statistically independent samples drawn from the master equation using some approximative sampling technique (e.g., Gaussian or Poisson leaping). In statistics, two-sample testing refers to a hypothesis testing problem that deals with the following null and alternative hypotheses:

H_0: $P = \widetilde{P}$

H_A: $P \neq \widetilde{P}$,

where P denotes the (unknown) solution $p_{\boldsymbol{x}}(x;t)$ of the master equation at time t from which the exact samples $\{x_1(t), x_2(t), \ldots, x_L(t)\}$ are drawn, whereas, $\widetilde{P}$ denotes the (unknown) distribution $\widetilde{P}(x;t)$ from which the approximate samples $\{\widetilde{x}_1(t), \widetilde{x}_2(t), \ldots, \widetilde{x}_L(t)\}$ are drawn. The TSKS test computes the *two-sample* Kolmogorov-Smirnov statistic:

$$S(t) = \max_x \left| G(x;t) - \widetilde{G}(x;t) \right|, \tag{4.4.22}$$

where G and $\widetilde{G}$ are the two empirical CDFs

$$G(x;t) = \frac{1}{L}\sum_{l=1}^{L}[x_l(t) \leq x] \quad \text{and} \quad \widetilde{G}(x;t) = \frac{1}{L}\sum_{l=1}^{L}[\widetilde{x}_l(t) \leq x]. \tag{4.4.23}$$

It is clear that the TSKS statistic is obtained from the KS statistic by replacing the CDF $\widehat{F}$ corresponding to the analytical approximation method with the empirical CDF $\widetilde{G}$ calculated from the data drawn from the master equation using the approximative sampling technique. Using an existing statistical software package, we can calculate a p-value for a given value of S. If this p-value is less than a chosen significance level α, then we can reject the validity of the approximate samples. This is due to the fact that, in this case, the difference between the empirical CDF of the approximate samples and the empirical CDF of the exact samples will be statistically significant. As a matter of fact, the p-value is less than α if and only if the TSKS statistic S is larger than a critical value $s_0(\alpha)$.

Similarly to KS testing, the $(1-\alpha)\%$ confidence interval for the true CDF $F(x;t)$ is given by (this is a consequence of Example 4.4.6 in Ref. [149] and Theorem 4 in Ref. [151])

$$\mathcal{C}_\alpha(x;t) = \big(\max\{0, G(x;t) - s_0(\alpha)\}, \min\{1, G(x;t) + s_0(\alpha)\}\big). \tag{4.4.24}$$

In this case, if the empirical CDF $\widetilde{G}(x;t)$ is not entirely within the confidence band $\mathcal{C}_\alpha(t) := \{\mathcal{C}_\alpha(x;t), -\infty < x < \infty\}$, then this approximation must be rejected, since $S(t) > s_0(\alpha)$ in this case. On the other hand, if $\widetilde{G}(x;t)$ lies entirely within $\mathcal{C}_\alpha(t)$, then we can claim with $(1-\alpha)\%$ confidence that the empirical CDF obtained by approximative sampling is close to the true CDF, in a sense quantified by the width $2s_0(\alpha)$ of the confidence band. We say in this case that we are at least $(1-\alpha)\%$ certain that the approximate samples provide an acceptable alternative to the exact samples. Note that we cannot derive a meaningful decision band in this case, since we do not know the exact probability distribution $\widetilde{P}$.

Unfortunately, it is not easy to determine the critical value of the TSKS test, due to the fact that the probability distribution $\widetilde{P}$ from which the approximative samples are drawn is discrete. However, this value is always smaller than the critical value obtained by assuming that $\widetilde{P}$

is continuous, given by [compare with Eq. (4.3.11)] [151]

$$s_0(\alpha) = \frac{\sqrt{|\ln(\alpha/2)/2|}}{\sqrt{L/2} + 0.12 + 0.11/\sqrt{L/2}} \,. \tag{4.4.25}$$

In this case, the maximum probability of the Type I error associated with the test that uses the critical value s_0 given by Eq. (4.4.25) will be no more than the maximum probability of the Type I error of the test that uses the true critical value.[7] However, the power of the test that uses critical value s_0 will be smaller than the power of the test that uses the true critical value, in which case we are dealing with a larger probability for the Type II error (i.e., a more conservative test). This is also reflected by the fact that the confidence band $\mathcal{C}_\alpha(t)$ will be wider when using s_0. Note however that $s_0 \sim (L)^{-1/2}$. As a consequence, we can reduce the width $2s_0$ of the confidence band and increase the power of the test by increasing the sample size L when this is possible.

Choosing the sample size and significance level

The problem of choosing values for L and α when testing the validity of an approximative sampling technique is similar to the problem of choosing these values when testing for the validity of an analytical approximation. However, only the two Scenarios 1 & 2, discussed in Section 4.3.1 are relevant here, since we no longer have analytical knowledge of the approximating probability distribution. TSKS testing requires twice as many samples as KS testing in order to obtain the same size confidence bands for the same level of significance. When sufficient computational resources are available, we can increase the power of the TSKS test if necessary by increasing the number of samples acquired by exact and approximative sampling. However, the fact that an investigator is considering approximative sampling may indicate that exact sampling is computationally demanding, perhaps due to a large and stiff chemical reaction system at hand. In this case, increasing the power of TSKS by increasing L may not be feasible, and the investigator may decide to use more powerful two-sample multivariate testing procedures at the expense of loosing the ability to compute confidence bands.

[7] This is a consequence of the fact that, since the true critical value is smaller than the critical value s_0, the significance level of the test (which equals the maximum probability of the Type I error) that uses the true critical value will be larger than the significance level of the test that uses s_0.

As we will demonstrate in Section 4.5, confidence bands can provide a wealth of useful information about the validity of an analytical approximation or an approximative sampling technique.

4.4.2 Extension to multiple dimensions

We can extend the TSKS test to multiple dimensions by using a similar procedure to the one discussed in Section 4.3.2. In particular, for each $n = 1, 2, \ldots, N$, we draw L independent vector samples $\{\boldsymbol{x}_l(t), l = 1, 2, \ldots, L\}$ from the master equation using exact sampling and L independent vector samples $\{\widetilde{\boldsymbol{x}}_l(t), l = 1, 2, \ldots, L\}$ using approximative sampling. We then compute the empirical marginal CDFs

$$G_n(x;t) = \frac{1}{L}\sum_{l=1}^{L}[x_{l,n}(t) \leq x] \quad \text{and} \quad \widetilde{G}_n(x;t) = \frac{1}{L}\sum_{l=1}^{L}[\widetilde{x}_{l,n}(t) \leq x], \tag{4.4.26}$$

and the two-sample KS statistic $S_n(t) = \max_x |G_n(x;t) - \widetilde{G}_n(x;t)|$. Subsequently, we use the TSKS hypothesis testing procedure to compute the p-value $p_n(t)$ and the confidence interval

$$\mathcal{C}_\alpha^n(x;t) = \big(\max\{0, G_n(x;t) - s_0(\alpha)\}, \min\{1, G_n(x;t) + s_0(\alpha)\}\big), \tag{4.4.27}$$

where $s_0(\alpha)$ is given by Eq. (4.4.25). After this process is complete, we combine the resulting p-values into one p-value $p(t)$, using Eq. (4.3.19), and reject the approximate samples obtained at time t with significance level α, whenever $p(t) < \alpha$.

4.5 Results

In this section, we illustrate the previous methods using two simple reaction models: the Schlögl model of chemistry and the SIR model of epidemiology, which can be used to model an autocatalytic reaction (e.g., autophosphorylation) coupled with an isomerization reaction. Although we could consider more complex systems, the two models discussed in this section are simple enough to allow numerical computation of the true solution of the master equation, while they are complex enough to provide a clear illustration of various concepts associated with our methods.

Example 1: The LNA method in the Schlögl model

It is a well-known fact that the Schlögl model [165] is a bistable chemical reaction system that can be described using a one-dimensional master equation [93]. It is also known that the LNA method can break down in a bistable system, producing variance estimates that can even diverge to infinity [5]. Therefore, the common advice regarding the LNA method is to avoid this approximation in systems whose underlying macroscopic equations have more than one stable fixed point. We demonstrate in this example that we may be able to use the LNA method to provide a useful analytical approximation to the solution of the master equation, even in bistable systems. Moreover, we show that the uncertainty regarding the validity of this approximation can be quantified.

The Schlögl model consists of the following four reactions:

$$\begin{aligned} 2X_1 + X_2 &\to 3X_1 \\ 3X_1 &\to 2X_1 + X_2 \\ X_1 &\to X_3 \\ X_3 &\to X_1, \end{aligned} \tag{4.5.28}$$

where the *concentrations* of X_2 and X_3 are held constant at levels γ_2 and γ_3, respectively. These reactions occur with mass-action propensity functions, given by

$$\begin{aligned} \pi_1(x_1) &= \frac{k_1}{V}\,\gamma_2\,x_1(x_1-1) \\ \pi_2(x_1) &= \frac{k_2}{V^2}\,x_1(x_1-1)(x_1-2) \\ \pi_3(x_1) &= k_3\,x_1 \\ \pi_4(x_1) &= k_4 V \gamma_3, \end{aligned} \tag{4.5.29}$$

where x_1 denotes the population number of X_1, k_1, k_2, k_3, k_4 are rate constants of the corresponding reactions, and V is the system volume.

By setting $\Omega = V$, we find that $\widetilde{X}_1(t) := X_1(t)/\Omega$ is the concentration of the chemical species X_1 in the system at time t. In this case, by examining Eq. (4.2.2) and the previous propensity functions, we find that $f(\Omega) = \Omega$, whereas $\widetilde{\pi}_1(y_1) \simeq k_1\gamma_2 y_1^2$, $\widetilde{\pi}_2(y_1) \simeq k_2 y_1^3$, $\widetilde{\pi}_3(y_1) = k_3 y_1$, and

$\widetilde{\pi}_4(y_1) = k_4\gamma_3$, for large enough V. Therefore, the macroscopic equation (4.2.3) is given by

$$\frac{d\mu_1(t)}{dt} = k_1\gamma_2\mu_1^2(t) - k_2\mu_1^3(t) - k_3\mu_1(t) + k_4\gamma_3, \tag{4.5.30}$$

with initial condition $\mu_1(0) = x_{0,1}/V$, where $\mu_1(t) = \mathrm{E}[\widetilde{X}_1(t)]$. Moreover, the Lyapunov equation (4.2.5) is given by

$$\frac{dc_{11}(t)}{dt} = k_1\gamma_2\mu_1^2(t) + k_2\mu_1^3(t) + k_3\mu_1(t) + k_4\gamma_3 + 2c_{11}(t)\left[2k_1\gamma_2\mu_1(t) - 3k_2\mu_1^2(t) - k_3\right], \tag{4.5.31}$$

with initial condition $c_{11}(0) = 0$.

We consider the following parameter values: $x_{0,1} = 600$ molecules, $\gamma_2 = 1$ molecule fl^{-1}, $\gamma_3 = 2$ molecules fl^{-1}, $V = 80$ fl, $k_1 = 3$ $\mathrm{fl}^2\mathrm{molecules}^{-3}\mathrm{hr}^{-1}$, $k_2 = 0.6$ $\mathrm{fl}^2\mathrm{molecules}^{-3}\mathrm{hr}^{-1}$, $k_3 = 2.95\,\mathrm{molecules}^{-1}\mathrm{hr}^{-1}$, and $k_4 = 0.25\,\mathrm{molecules}^{-1}\mathrm{hr}^{-1}$. These values are known to produce a bistable macroscopic system with stable fixed points at $\mu^* = 299.7827$ and $\mu^{**} = 17.1295$ [93]. We are interested in the following question:

> *We are running the Schlögl model for one hour. Will the LNA method accurately describe the solution to the underlying master equation at the end of this run?*

To answer this question, we begin by solving Eq. (4.5.30) using the "ode23s" numerical solver of MATLAB® and find that the LNA method predicts that the mean value of the population process $X_1(1)$ at the one hour mark equals 300.0157 molecules, when the system is initialized with $x_{0,1} = 600\,\mathrm{molecules}$, which is in the basin of attraction of $\Omega\mu^*$. We subsequently solve the Lyapunov equation (4.5.31) numerically and find that the standard deviation of the population process $X_1(1)$ equals 24.3197 molecules. Therefore, the LNA method predicts that the solution of the master equation at the one hour mark is given by

$$\widehat{P}^*(x;1) = 0.0164\exp\Big\{-\frac{(x-300.0157)^2}{1182.8934}\Big\}. \tag{4.5.32}$$

Our goal is to statistically verify the validity of the LNA method when the system is initialized with $x_{0,1} = 600$ molecules, using KS testing.

As we have discussed earlier, KS testing requires specification of the sample size L and significance level α. Note that this system fits well with our discussion pertaining to Scenario 3 in Section 4.3.1. We therefore proceed by using the algorithm discussed in Appendix C.

To find appropriate values for L and α, we focus on satisfying the following two requirements. When we are at least 99% certain that the master equation is initialized within the basin of attraction of the first stable fixed point $\Omega\mu^*$, we can accept a test with less than 5% power. Moreover, when we are at most 90% certain that the master equation is initialized within the basin of attraction of the second stable fixed point $\Omega\mu^{**}$, we must accept a test with at least 95% power. This implies the following values for the parameters associated with the algorithm discussed in Appendix C: $r_0 = 0.9$, $r_1 = 0.99$, $b_0 = 0.95$, and $b_1 = 0.05$. We must also find the probability distribution $\widehat{P}^{**}$, which describes the behavior of the LNA method within the basin of attraction of $\Omega\mu^{**}$. In this case, the LNA method predicts that the mean tends towards $\Omega\mu^{**} = 17.1295$, whereas, the standard deviation tends toward 17.1295 molecules. Therefore, we set

$$\widehat{P}^{**}(x;1) = 0.0739 \exp\Big\{ - \frac{(x - 17.1295)^2}{58.3632} \Big\}, \tag{4.5.33}$$

which fully specifies all information needed by the algorithm discussed in Appendix C. This algorithm results in $L = 350$ and $\alpha = 0.025$, from which we obtain $s_0(\alpha) = 0.0786$ by means of Eq. (4.3.11).

In Fig. 4.1(a), we depict a portion of the decision boundaries of the resulting decision band $\mathcal{D}_\alpha(1)$ (red lines), whose width is $2s_0(\alpha) = 0.1572$, together with the CDF $\widehat{F}(x;1)$ obtained by the LNA method (blue line). If we are satisfied with this decision band, we can proceed to perform KS testing. After drawing $L = 350$ samples from the master equation using exact sampling and after computing the empirical CDF $G(x;1)$ (gray line), we find that the test fails to reject the LNA method, since the value of the KS statistic is $S(1) = 0.0759$, which is less than the critical value $s_0(\alpha) = 0.0786$. This is also true since all values of the empirical CDF $G(x;1)$, depicted by the gray line in Fig. 4.1(a), lie inside $\mathcal{D}_\alpha(1)$. In Fig. 4.1(b), we depict a portion of the boundaries of the $(1-\alpha)\% = 97.5\%$ confidence band $\mathcal{C}_\alpha(1)$ produced by the KS test (orange lines), together with

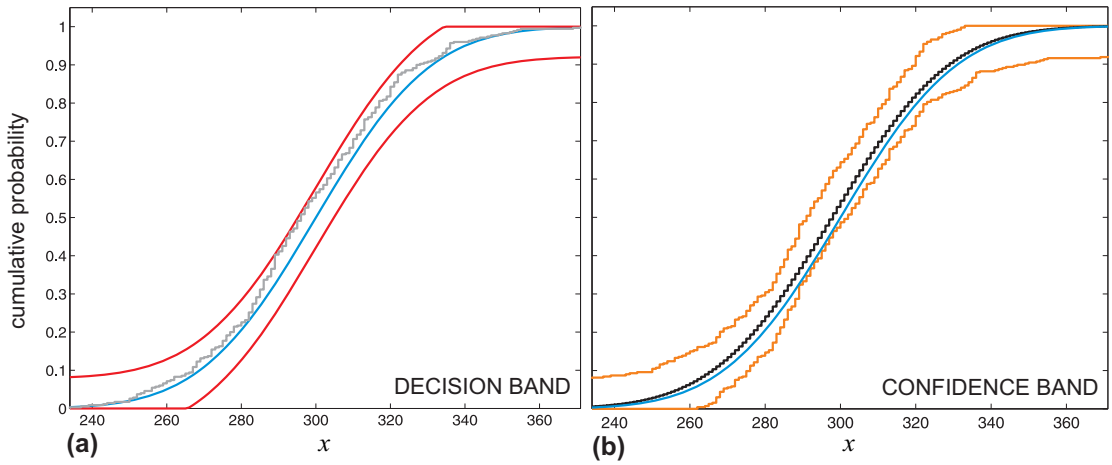

Figure 4.1: (a) Portion of the decision band $\mathcal{D}_\alpha(1)$ of the KS test obtained for the Schlögl model, with $L = 350$ and $\alpha = 0.025$. The red lines depict the decision boundaries, whereas, the blue line depicts the CDF $\widehat{F}(x;1)$ of $X_1(1)$ obtained by the LNA method. The values of the empirical CDF $G(x;1)$, depicted by the gray line, computed from L samples drawn from the master equation by exact sampling, lie inside $\mathcal{D}_\alpha(1)$. Therefore, KS hypothesis testing fails to reject the LNA method. (b) Portion of the corresponding 97.5% confidence band $\mathcal{C}_\alpha(1)$. The orange lines depict the confidence boundaries, whereas, the blue line depicts the CDF $\widehat{F}(x;1)$ of $X_1(1)$ obtained by the LNA method. The black line depicts the true CDF $F(x;1)$ numerically obtained with the KSA method.

the LNA CDF $\widehat{F}(x;1)$ (blue line). The confidence band predicts (with 97.5% certainty) that the true CDF $F(x;1)$ (black line) will be contained entirely within $\mathcal{C}_\alpha(1)$ (orange lines), and Fig. 4.1(b) demonstrates the accuracy of this prediction.

Note that the LNA CDF is closer to the lower boundary of the confidence region than the upper boundary. This tells us that the probability $\Pr[X_1(1) \leq x]$ predicted by the LNA method is very likely to underestimate the true probability. This does not come as a surprise, since we understand that the LNA method is ignoring the existence of the second stable fixed point at $\Omega\mu^{**}$ and thus the probability mass assigned to the vicinity of that point. It is interesting to note that the confidence band depicted in Fig. 4.1(b) could have warned an unsuspecting user of the LNA method of this possibility.

The previous results clearly indicate that the LNA method provides an acceptable analytical approximation to the master equation at the one hour mark. Moreover, Fig. 4.1(b) provides a quantitative assessment of the quality of the approximation, which helps us decide whether we can

trust this approximation. To confirm these results, we solved the master equation numerically using the KSA method.[8] The true CDF, depicted by the black line in Fig. 4.1(b), falls right in the middle of the confidence region and closely agrees with the LNA CDF (blue line). To measure the difference between the corresponding probability density functions, we computed the Kullback-Leibler distance [126]. This produced a value of 0.0072 bits, which indicates that there is almost no loss of information when characterizing the true solution of the master equation using the approximation obtained by the LNA method.

As a final note, we should point out another advantage of calculating and using confidence bands. As indicated in Fig. 4.1(b), the exact solution to the master equation does differ slightly from the solution obtained by the LNA method. For large enough L, these differences will be picked up by the test, which will reject the approximate solution obtained by the LNA method. Indeed, the p-value obtained in the current example (which equals 0.0337) was slightly above the significance level $\alpha = 0.025$, indicating that the test almost rejected the LNA method due to its proximity to the lower boundary of the confidence band. As $L \to \infty$, the confidence boundaries will move towards the center of the confidence band. In this case, the test will become increasingly more certain of what the true solution is and it will start rejecting the LNA method. This however is not a problem in practice, since we can quantitatively observe how the LNA CDF strays from the confidence boundaries and make a decision, with a certain level of confidence, as to whether or not the LNA method produces an acceptable approximation to the solution of the master equation. This discussion demonstrates the fact that the raw outcome of the test (i.e., reject or fail to reject the LNA method) is less informative than the confidence bands produced by the test.

Example 2: The LNA method in the SIR model

The SIR model of epidemiology consists of the following two reactions:

$$\begin{aligned} X_1 + X_2 &\to 2X_2 \\ X_2 &\to X_3. \end{aligned} \tag{4.5.34}$$

[8] After applying a finite projection step to reduce the size of the underlying state-space [30].

The first reaction models infection of an individual X_1, who is susceptible to a disease (such as the flu), by an infected individual X_2, whereas, the second reaction models the recovery of an infected individual X_2. These reactions occur with mass-action propensity functions, given by

$$\begin{aligned} \pi_1(x_1, x_2) &= \frac{k_1}{I}\, x_1 x_2 \\ \pi_2(x_1, x_2) &= k_2 x_2, \end{aligned} \tag{4.5.35}$$

where x_1, x_2 denote the numbers of X_1 and X_2 individuals, respectively, k_1, k_2 are the rate constants of the corresponding reactions, and I denotes the total number of individuals in the system. Although the SIR model appears to be three-dimensional, the constraint $X_1(t) + X_2(t) + X_3(t) = I$ means that the system is in fact two-dimensional.

By setting $\Omega = I$, we find that $\widetilde{X}_1(t) := X_1(t)/\Omega$, $\widetilde{X}_2(t) := X_2(t)/\Omega$ are respectively the fractions of the susceptible and infected individuals X_1 and X_2 in the system at time t. In this case, by examining Eq. (4.2.2) and the previous propensity functions, we find that $f(\Omega) = \Omega$, whereas $\widetilde{\pi}_1(y_1, y_2) = k_1 y_1 y_2$ and $\widetilde{\pi}_2(y_1, y_2) = k_2 y_2$. Therefore, the macroscopic equations (4.2.3) are given by

$$\begin{aligned} \frac{d\mu_1(t)}{dt} &= -k_1 \mu_1(t)\mu_2(t) \\ \frac{d\mu_2(t)}{dt} &= k_1 \mu_1(t)\mu_2(t) - k_2 \mu_2(t), \end{aligned} \tag{4.5.36}$$

initialized by $\mu_1(0) = x_{0,1}/I$ and $\mu_2(0) = x_{0,2}/I$, where $\mu_1(t) = \mathrm{E}[\widetilde{X}_1(t)]$ and $\mu_2(t) = \mathrm{E}[\widetilde{X}_2(t)]$. Moreover, the Lyapunov equations (4.2.5) are given by

$$\begin{aligned} \frac{dc_{11}(t)}{dt} &= -2k_1\mu_2(t)c_{11}(t) - 2k_1\mu_1(t)c_{12}(t) + k_1\mu_1(t)\mu_2(t) \\ \frac{dc_{12}(t)}{dt} &= -k_1\mu_1(t)c_{22}(t) + k_1\mu_2(t)c_{11}(t) + \big[k_1\mu_1(t) - k_2 - k_1\mu_2(t)\big]c_{12}(t) - k_1\mu_1(t)\mu_2(t) \\ \frac{dc_{22}(t)}{dt} &= 2k_1\mu_2(t)c_{12}(t) + 2\big[k_1\mu_1(t) - k_2\big]c_{22}(t) + k_1\mu_1(t)\mu_2(t) + k_2\mu_2(t), \end{aligned} \tag{4.5.37}$$

with initial condition $c_{11}(0) = c_{12}(0) = c_{22}(0) = 0$.

We consider the following parameter values: $I = 763$, $x_{0,1} = 762$, $x_{0,2} = 1$, $k_1 = 1.6633$ individuals^{-1}day^{-1}, and $k_2 = 0.44036$ individuals^{-1}day^{-1}. These values have been esti-

mated based on real data obtained from a 1978 flu outbreak in an English boarding school [135,136]. Similarly to the previous example, we are interested in the following question:

> *We are running the SIR model for six days. Will the LNA method accurately describe the solution to the underlying master equation at the end of this run?*

We chose to simulate the model for six days since the peak infection occurs at the six day mark according to the available data [135].

To answer the previous question, we begin by solving the system of macroscopic equations (4.5.36) using the "ode23s" numerical solver of MATLAB® and find that the LNA method predicts mean values 253.1923 and 287.5587 for the population processes $X_1(6)$ and $X_2(6)$, respectively, at the six day mark. We subsequently solve the Lyapunov equations (4.5.37) numerically and find that $c_{11}(6) = 40.4421$, $c_{12}(6) = -9.2960$, and $c_{22}(6) = 2.5231$. These values fully specify the Gaussian approximation to the master equation associated with the SIR model derived by the LNA method. Our goal is to statistically verify the validity of this approximation using KS testing.

To specify the required sample size L and significance level α of the KS test, we follow Scenario 1 in Section 4.3.1. We limit the probability of Type I error to 1% and the width $2s_0(\alpha)$ of the confidence band to size $w_0 = 0.1$. In this case, $\alpha = 0.01$, whereas $L = 1{,}060$ by taking the lowest possible value suggested by Eq. (4.3.15). Moreover, Tippett's method results in $\alpha_0 = 0.005$.

In Fig. 4.2(a), we depict the decision boundaries of the resulting decision bands $\mathcal{D}^1_{\alpha_0}(6)$ and $\mathcal{D}^2_{\alpha_0}(6)$ (red lines) together with the marginal CDFs $\widehat{F}_1(x;6)$ and $\widehat{F}_2(x;6)$, obtained by the LNA method (blue lines). The width of these decision bands is $2s_0(\alpha_0) = 0.1058$. To perform KS testing, we draw two independent sets of $L = 1{,}060$ samples from the master equation using exact sampling and use these samples to compute the empirical marginal CDFs $G_1(x;6)$ and $G_2(x;6)$ (gray lines). We find that KS hypothesis testing based on Tippett's method rejects the LNA method. This is also clear from Fig. 4.2(a), since the computed empirical marginal CDFs fall outside the corresponding decision regions. In Fig. 4.2(b), we depict the boundaries of the $(1-\alpha)\% = 99\%$ confidence bands $\mathcal{C}^1_\alpha(6)$ and $\mathcal{C}^2_\alpha(6)$ produced by the KS test (orange lines), together with the marginal LNA CDFs

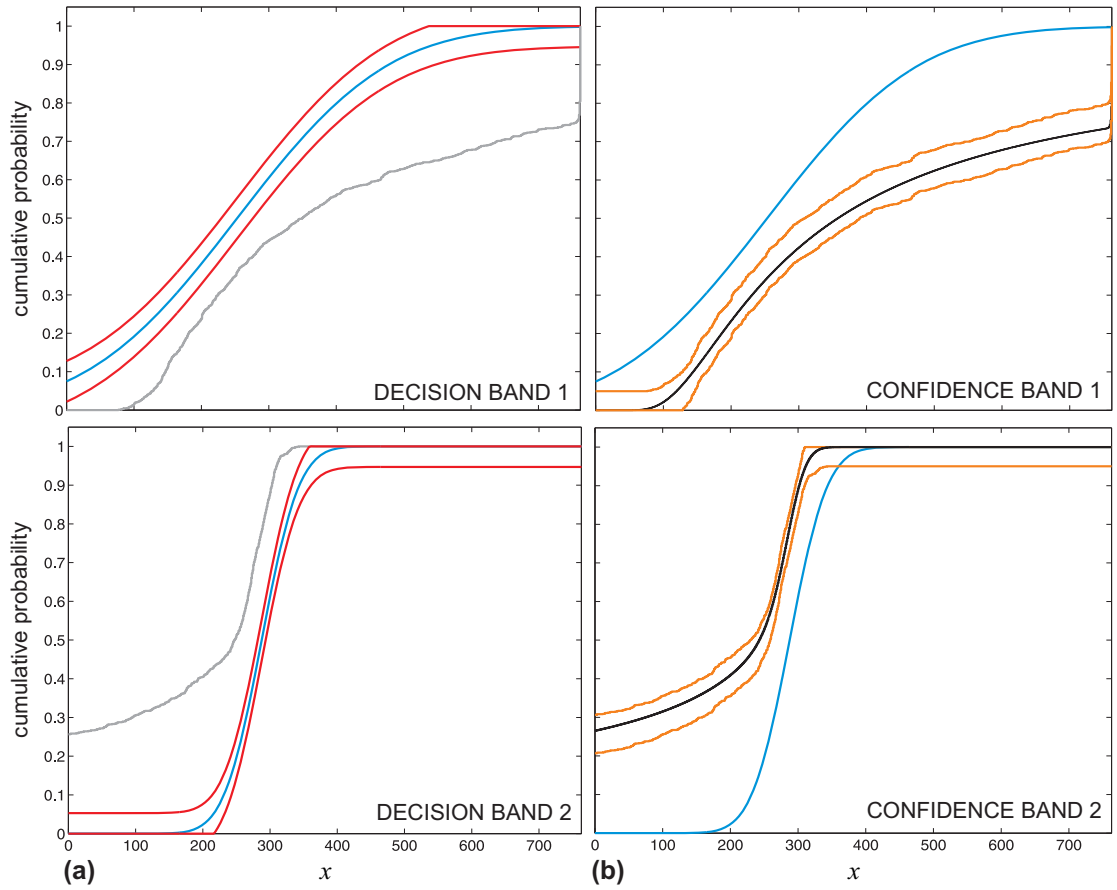

Figure 4.2: (a) The decision bands $\mathcal{D}^1_{\alpha_0}(6)$ and $\mathcal{D}^2_{\alpha_0}(6)$ of the KS test obtained for the SIR model, with $L = 1{,}060$ and $\alpha = 0.01$. The red lines depict the decision boundaries, whereas, the blue lines depict the CDFs $\widehat{F}_1(x;6)$ and $\widehat{F}_2(x;6)$ of $X_1(6)$ and $X_2(6)$, respectively, obtained by the LNA method. Note that the empirical marginal CDFs $G_1(x;6)$ and $G_2(x;6)$, depicted by the gray lines, computed from $2L$ samples drawn from the master equation by exact sampling, lie outside the corresponding decision bands. Therefore, KS hypothesis testing rejects the LNA method. (b) The 99% confidence bands $\mathcal{C}^1_\alpha(6)$ and $\mathcal{C}^2_\alpha(6)$ of the KS test obtained for the SIR model. The orange lines depict the confidence boundaries, whereas, the blue lines depict the CDFs $\widehat{F}_1(x;6)$ and $\widehat{F}_2(x;6)$ of $X_1(6)$ and $X_2(6)$, respectively, obtained by the LNA method. The black lines depict the true marginal CDFs $F_1(x;6)$ and $F_2(x;6)$, numerically obtained with the IE method.

$\widehat{F}_1(x;6)$ and $\widehat{F}_2(x;6)$ (blue lines). The width of these bands is $2s_0(\alpha) = 0.0996$. We expect the true marginal CDFs $F_1(x;6)$ and $F_2(x;6)$ (black lines) to be inside the corresponding confidence bands with 99% probability.

Clearly, the marginal CDFs obtained by the LNA method lie outside the confidence bands, indicating that this method provides a poor approximation to the solution of the master equation. This is in direct contrast to a relatively recent attempt to approximately solve the SIR model pre-

sented here using the LNA method [17]. In fact, the KS test produces extraordinarily small p-values, $p_1(6) = 5.0962 \times 10^{-87}$ and $p_2(6) = 5.7225 \times 10^{-138}$, which lead to a combined p-value of zero (within numerical precision). Therefore, the KS test will reject the LNA method at any level of significance. It turns out that the Kullback-Leibler distance between the probability distributions obtained by the IE and LNA methods equals 27.5953 bits. This indicates that there is significant information loss when using the approximation obtained by the LNA method, confirming the importance of the test's decision to reject the approximation.

The previous example clearly demonstrates the practical value of the proposed hypothesis testing approach. Without this technique, we are left to decide whether or not to use the LNA method based on limiting arguments (as $\Omega \to \infty$), which are of little or no help in practice. KS hypothesis testing is an effective way to deal with this problem and can be used to provide useful information on when and why the LNA method breaks down. For example, the confidence band $\mathcal{C}^1_\alpha(6)$ depicted in Fig. 4.2(b) indicates, with 99% certainty, a significant probability that no susceptible individual will get infected during a six day period [the probability of $X_1(6) = 762$ is between 0.1 and 0.2]. This for example may happen if the flu virus, after infecting one individual, dies before infecting anybody else. Moreover, the confidence band $\mathcal{C}^2_\alpha(6)$ depicted in Fig. 4.2(b) indicates, with 99% certainty, a significant probability that there will be no infected individuals at day 6 [the probability of $X_2(6) = 0$ is between 0.2 and 0.3], which can happen for example if all infected individuals recover by day 6. Note however that both of these facts are not predicted by the approximation obtained using the LNA method, which wrongly assigns zero probability to having 762 susceptible and 0 infected individuals at day 6.

Example 3: Gaussian leaping in the SIR model

Since we found that the LNA method provides an unacceptable approximation to the solution of the master equation associated with the SIR model, we may decide to study marginal statistical properties of the model using Monte Carlo analysis based on approximative sampling. We first start with Gaussian leaping, due to its computational efficiency. As with many approximation

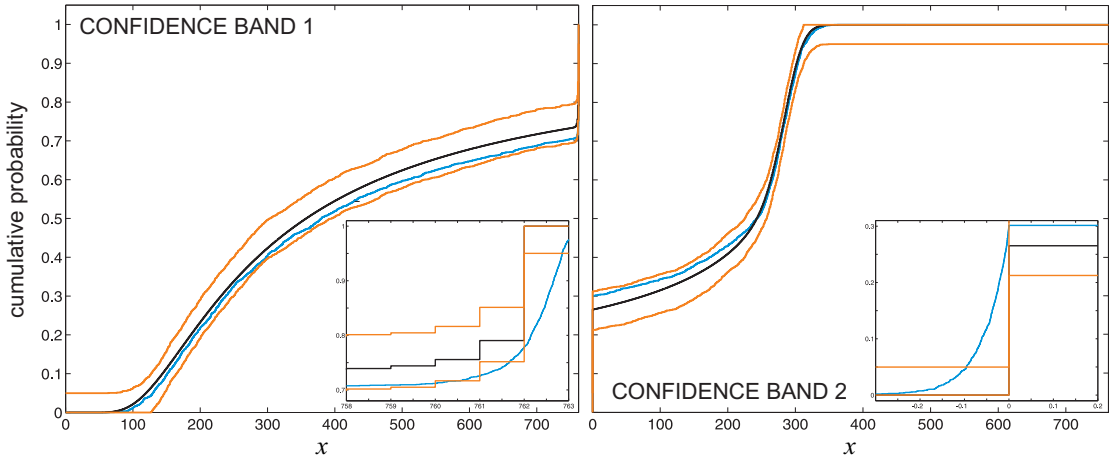

Figure 4.3: The 99% confidence bands $\mathcal{C}_\alpha^1(6)$ and $\mathcal{C}_\alpha^2(6)$ of the TSKS test obtained for the SIR model, with $L = 2{,}120$ and $\alpha = 0.01$. The orange lines depict the confidence boundaries, whereas, the blue lines depict the empirical CDFs $\widetilde{G}_1(x;6)$ and $\widetilde{G}_2(x;6)$ of $X_1(6)$ and $X_2(6)$, respectively, obtained by approximative sampling using Gaussian leaping. The black lines depict the true marginal CDFs $F_1(x;6)$ and $F_2(x;6)$, numerically obtained with the IE method. The marginal empirical CDFs obtained by Gaussian leaping remain within the confidence bands, except at values close to the left and right boundaries of the state-space (insets). TSKS hypothesis testing rejects the validity of the approximative samples in this case.

algorithms, the validity of Gaussian leaping depends on the choice of the leaping parameter τ, which must meet certain criteria that are often difficult or impossible to verify. We are thus interested in the following question:

> *We are running the SIR model for six days. Will Gaussian leaping with $\tau = 0.05$ produce population samples at day 6 that can safely replace, in a Monte Carlo study of marginal statistics, samples obtained by exact sampling?*

To answer this question using hypothesis testing, we begin by limiting as before the probability of Type I error to 1% and the width $2s_0(\alpha)$ of the confidence band to size $w_0 = 0.1$. In this case, $\alpha = 0.01$, whereas, Eq. (4.3.15) and Eq. (4.4.25) imply that $L = 2{,}120$. Subsequently, we independently draw two sets of L samples from the master equation by employing exact sampling and Gaussian leaping with $\tau = 0.05$ days, we use these samples to calculate the marginal empirical CDFs $G_1(x;6)$, $G_2(x;6)$ and $\widetilde{G}_1(x;6)$, $\widetilde{G}_2(x;6)$, and we finally proceed with hypothesis testing. It turns out that TSKS hypothesis testing based on Tippett's method rejects the samples obtained by Gaussian leaping with zero p-value.

In Fig. 4.3, we depict the boundaries of the resulting $(1-\alpha)\% = 99\%$ confidence bands $\mathcal{C}^1_\alpha(6)$ and $\mathcal{C}^2_\alpha(6)$ (orange lines) together with the marginal empirical CDFs $\widetilde{G}_1(x;6)$ and $\widetilde{G}_2(x;6)$ computed from the approximative samples obtained by Gaussian leaping (blue lines). The width of these confidence bands is $2s_0(\alpha) = 0.0996$. We expect the true marginal CDFs $F_1(x;6)$ and $F_2(x;6)$ (black lines) to be inside the corresponding confidence bands with 99% probability.

These results clearly indicate that Gaussian leaping provides a fairly acceptable approximation to the CDFs over most values of x. However, they also point out exactly where and how Gaussian leaping breaks down. It is clear from the insets depicted in Fig. 4.3 that Gaussian leaping produces an empirical CDF $\widetilde{G}_1(x;6)$ whose values at population levels close to 762 may lie outside the confidence band $\mathcal{C}^1_\alpha(6)$. Although this confidence band suggests, with 99% certainty, that there is very small probability [at most $s_0(\alpha) = 0.0498$] to have more than 762 susceptible individuals at day 6, the empirical CDF $\widetilde{G}_1(x;6)$ predicts significant probability for this impossible event.[9] Likewise, Gaussian leaping produces an empirical CDF $\widetilde{G}_2(x;6)$ whose values at population levels below 0 may lie outside the confidence band $\mathcal{C}^2_\alpha(6)$. Although this confidence band suggests, with 99% certainty, that there is very small probability [at most $s_0(\alpha) = 0.0498$] to have less than 0 infected individuals at day 6, the empirical CDF $\widetilde{G}_2(x;6)$ wrongly predicts significant probability for this impossible event (clearly negative populations are meaningless). The TSKS test picks up these flaws returning extremely small p-values $p_1(6) = 5.0669 \times 10^{-46}$ and $p_2(6) = 1.0445 \times 10^{-84}$. As a consequence, Tippett's method produces a combined p-value that (within numerical precision) equals zero. This results in rejecting the samples obtained by Gaussian leaping at any level of significance.

Example 4: Poisson leaping in the SIR model

Given that Gaussian leaping is an unacceptable sampling scheme for the SIR model, we would now like to investigate whether the more accurate Poisson leaping algorithm with the same value of τ can produce statistically acceptable samples for Monte Carlo analysis. We are therefore interested in the following question:

[9] The initial number of susceptible individuals is 762, Hence, it is not possible to have more than 762 susceptible individuals in the system.

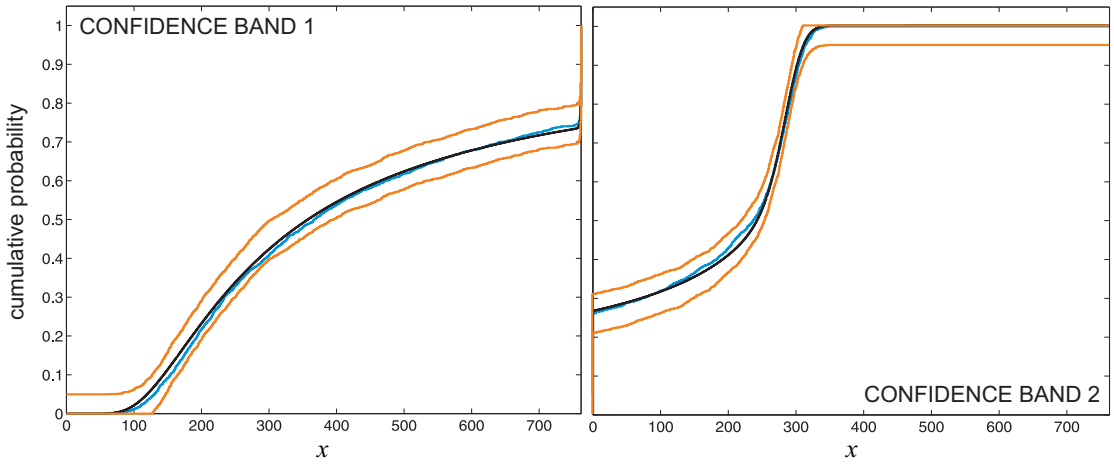

Figure 4.4: The 99% confidence bands $\mathcal{C}_\alpha^1(6)$ and $\mathcal{C}_\alpha^2(6)$ of the TSKS test obtained for the SIR model, with $L = 2{,}120$ and $\alpha = 0.01$. The orange lines depict the confidence boundaries, whereas, the blue lines depict the empirical CDFs $\widetilde{G}_1(x;6)$ and $\widetilde{G}_2(x;6)$ of $X_1(6)$ and $X_2(6)$, respectively, obtained by approximative sampling using Poisson leaping with $\tau = 0.05$. The black lines depict the true marginal CDFs $F_1(x;6)$ and $F_2(x;6)$, numerically obtained with the IE method. TSKS hypothesis testing fails to reject the validity of the approximative samples in this case. Since the marginal empirical CDFs obtained by Poisson leaping remain within the $(1-\alpha)\%$ confidence bands, we can accept the approximative samples with 99% confidence.

> *We are running the SIR model for six days. Will Poisson leaping with $\tau = 0.05$ produce population samples at day 6 that can safely replace, in a Monte Carlo study of marginal statistics, samples obtained by exact sampling?*

To answer this question using hypothesis testing, we begin by limiting as before the probability of Type I error to 1% and the width $2s_0(\alpha)$ of the confidence band to size $w_0 = 0.1$. In this case, $\alpha = 0.01$, whereas, Eq. (4.3.15) and Eq. (4.4.25) imply that $L = 2{,}120$. Subsequently, we independently draw two sets of L samples from the master equation by employing exact sampling and Poisson leaping with $\tau = 0.05$ days, we use these samples to calculate the marginal empirical CDFs $G_1(x;6)$, $G_2(x;6)$ and $\widetilde{G}_1(x;6)$, $\widetilde{G}_2(x;6)$, and finally we proceed with hypothesis testing. It turns out that, in this case, TSKS hypothesis testing based on Tippett's method fails to reject the samples obtained by Poisson leaping.

In Fig. 4.4, we depict the boundaries of the resulting $(1-\alpha)\% = 99\%$ confidence bands $\mathcal{C}_\alpha^1(6)$ and $\mathcal{C}_\alpha^2(6)$ (orange lines) together with the marginal empirical CDFs $\widetilde{G}_1(x;6)$ and $\widetilde{G}_2(x;6)$ computed from the approximative samples obtained by Poisson leaping (blue lines). The width of

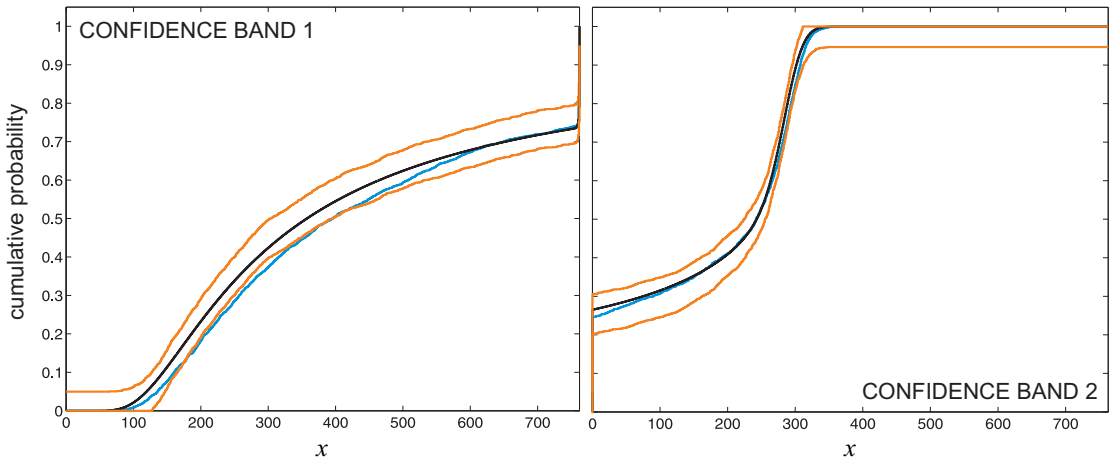

Figure 4.5: The 99% confidence bands $\mathcal{C}_\alpha^1(6)$ and $\mathcal{C}_\alpha^2(6)$ of the TSKS test obtained for the SIR model, with $L = 2{,}120$ and $\alpha = 0.01$. The orange lines depict the confidence boundaries, whereas, the blue lines depict the empirical CDFs $\widetilde{G}_1(x;6)$ and $\widetilde{G}_2(x;6)$ of $X_1(6)$ and $X_2(6)$, respectively, obtained by approximative sampling using Poisson leaping with $\tau = 0.08$. The black lines depict the true marginal CDFs $F_1(x;6)$ and $F_2(x;6)$, numerically obtained with the IE method. TSKS hypothesis testing rejects the validity of the approximative samples obtained in this case.

these confidence bands is again $2s_0(\alpha) = 0.0996$ and we expect the true marginal CDFs $F_1(x;6)$, $F_2(x;6)$ (black lines) to be inside the corresponding confidence bands with 99% probability. Note that both marginal empirical CDFs are within the corresponding confidence bands and do not experience the same boundary issues identified in the case of Gaussian leaping. Therefore, we may accept the samples obtained by Poisson leaping with 99% confidence. As a matter of fact, the TSKS test returned p-values $p_1(6) = 0.0645$ and $p_2(6) = 0.5415$, whereas, Tippett's method produced a combined p-value $p(6) = 0.1249$, which is larger than the significance level $\alpha = 0.01$. As a consequence, when $\tau = 0.05$ days, hypothesis testing fails to reject the samples obtained by Poisson leaping at 1% level of significance.

On the other hand, we might have asked the slightly different question:

> *We are running the SIR model for six days. Will Poisson leaping with $\tau = 0.08$ produce population samples at day 6 that can safely replace, in a Monte Carlo study of marginal statistics, samples obtained by exact sampling?*

We depict in Fig. 4.5 the boundaries of the resulting 99% confidence bands $\mathcal{C}_\alpha^1(6)$ and $\mathcal{C}_\alpha^2(6)$ (orange lines) together with the marginal empirical CDFs $\widetilde{G}_1(x;6)$ and $\widetilde{G}_2(x;6)$ computed from approximative samples obtained by Poisson leaping with the larger $\tau = 0.08$ leaping value (blue lines).

Although the marginal empirical CDF $\widetilde{G}_2(x;6)$ associated with the infected individuals is within the corresponding confidence band, values of the marginal empirical CDF $\widetilde{G}_1(x;6)$, associated with the susceptible individuals, fail to satisfy this property. As a matter of fact, for many values of x, $\widetilde{G}_1(x;6)$ is very close or below the lower confidence boundary of $\mathcal{C}^1_\alpha(6)$, which indicates that it underestimates the true CDF values. This is due to the fact that the required leaping condition – all propensity functions must be constant during any time interval $[j\tau,(j+1)\tau)$ – is not satisfied in this case. Although it is not in general possible to check whether or not the leaping condition is satisfied, the TSKS hypothesis test is capable of detecting that something is wrong and reject the approximative samples. In our case, the TSKS test returned p-values $p_1(6) = 1.6338 \times 10^{-5}$ and $p_2(6) = 0.0426$, whereas, Tippett's method produced a combined p-value $p(6) = 3.2674 \times 10^{-5}$, which is appreciably smaller than the significance level $\alpha = 0.01$. As a consequence, when $\tau = 0.08$ days, hypothesis testing rejects the samples obtained by Poisson leaping.

4.6 Discussion

The hypothesis testing framework proposed in this chapter provides a rigorous quantitative methodology for checking the validity of an approximation technique for solving the master equation. We can use KS hypothesis testing to reject or accept (with a certain level of confidence) the validity of an analytical approximation to the solution of the master equation or use TSKS testing to reject or accept (with a certain level of confidence) the validity of samples drawn from the master equation using an approximative sampling technique. Although substantial computational effort may be necessary to draw the samples required to implement the proposed methods, we believe that this effort is worthwhile. If hypothesis testing rejects the validity of a given approximation technique, then we may try to use or develop an alternative approximation method, instead of proceeding with analysis that might lead to erroneous results. On the other hand, if we can accept an approximation technique with a high level of confidence, then we can use it to replace exact sampling with appreciable

computational savings.

The proposed method requires that we draw exact samples from the master equation. However, we expect that the number of samples required for meaningful hypothesis testing to be appreciably less than the number of samples required for estimating the solution of the master equation via Monte Carlo. For instance, the state-space of the example considered in Section 4.5, which involves two "independent" species, includes $762 \times 763 = 581{,}406$ elements. Estimating the solution of the master equation over such a large state space using Monte Carlo would require a very large number of exact Monte Carlo samples. On the other hand, by drawing only $2 \times 1{,}060 = 2{,}120$ exact samples, we were able to decisively reject the LNA method.

In general, it is well-known that the "curse of dimensionality" demands an exponential increase in the number of samples required for sufficiently estimating the solution of the master equation via Monte Carlo. However, the number of samples required by the proposed hypothesis testing procedure increases only linearly with respect to the number of "independent" species in the reaction system. Therefore, we expect that the method can be effectively used in practice, provided that sufficient computational power is available.

An issue that deserves special attention is extension of the basic ideas presented in this chapter to a multivariate setting. To preserve the simplicity of KS hypothesis testing and the ability to compute decision and confidence bands, we decided to focus our effort on developing statistical testing methods applied on marginal distributions. If the marginal approach to hypothesis testing rejects the null hypothesis, this result immediately applies to the joint multivariate setting (e.g., since rejecting a marginal distribution forces us to reject the entire multivariate distribution). However, if the test fails to reject a marginal distribution, this does not allow us to reach a strong statistical conclusion about the validity of approximation methods when a multivariate statistical analysis of the master equation is of interest. As a consequence, it is important to develop a multivariate hypothesis testing approach that can address the problems discussed in this chapter in those practical situations that require multivariate statistical analysis of the master equation.

Although the methods discussed in this chapter are not capable of providing exact assurances for the validity of an approximation technique in a multivariate setting, from a practical perspective they do provide some additional empirical evidence, in conjunction to existing theoretical justifications, that the approximation can produce reasonable results. Knowledge of marginal distributions does not determine the multivariate distribution, but it appreciably reduces the space of all possible distributions. We believe that until effective multivariate hypothesis testing techniques become available, the current marginal techniques could be utilized even in a joint setting, if only to rule out approximations that produce poor results.

Our discussion in this chapter has been limited to the "static" problem of statistically checking the validity of an approximation method when a parameter of interest, such as time, system size, or the τ parameter in Gaussian or Poisson leaping, takes a specific value. We addressed this problem by employing a *single* hypothesis testing approach. An equally important problem however is checking the validity of an approximation method as a function of changing parameter values. For example, we may want to determine the minimum value of the system size Ω at which the LNA method fails to provide an acceptable solution to the master equation or adjudicate the maximum τ value for which Poisson leaping produces acceptable approximative sampling.

A naïve solution to the previous "dynamic" problem is to discretize the parameter of interest into Q values, independently apply KS or TSKS hypothesis testing for each value, and determine the extreme value at which the test rejects the approximative method. By doing so however we may end up with a test whose probability of Type I error (falsely rejecting the approximative method) within the batch of individual tests rapidly increases with Q, thus producing misleading results. For example, if $Q = 15$ and $\alpha = 0.05$ for each individual test, then the probability of observing at least one Type I error in a batch of Q tests is $1 - (1 - \alpha)^Q \simeq 0.54$, indicating that we are more likely than not to observe an error in this "dynamic" testing scenario.

On the other hand, we can develop a rigorous approach to the "dynamic" problem based on a *multiple* hypothesis testing procedure that is appropriately designed to control the probability

of one or more false rejections, known as family-wise error rate [148], or the expected proportion of falsely rejected hypotheses, known as false discovery rate [166]. If sufficient computational power is available, we can adapt the techniques proposed in this chapter to multiple hypothesis testing in a rather straightforward manner. For example, we can independently apply KS testing for each parameter value, compute the resulting p-values p_q, $q = 1, 2, \ldots, Q$, and order them in an increasing order $p_{[1]} \leq p_{[2]} \leq \cdots \leq p_{[Q]}$. If $H_0^{[q]}$ is the null hypothesis of rejecting the approximative technique at parameter value $[q]$, we may reject all null hypotheses $H_0^{[q]}$, for $q = 1, 2, \ldots, q^*$, where q^* is the largest q for which $p_{[q]} \leq (q/Q)\alpha$, with α being the significance level of each individual test. We may then be able from this result to determine the extreme value of q at which the test rejects the approximative method.

While the previous method, known as Bonferroni multiple-testing procedure, is straightforward, it can be computationally demanding. Moreover, it may turn out that this technique is not the most appropriate multiple hypothesis testing method for our current framework. Therefore, more research must be done on developing an effective and computationally efficient method for statistically testing the validity of an approximative technique for solving the master equation in a "dynamic" fashion.

Chapter 5

Thermodynamic Analysis of Leaky Markovian Networks

Up to this point, we have focused on (approximately or exactly) solving the master equation of Markov processes on networks. In this chapter, we take a different approach which focuses instead on thermodynamic potentials and averages that provide a far more compact system description. To demonstrate the power of the proposed approach, we consider the complex dynamical phenomenon of avalanching.

In this chapter, we specifically study the role intrinsic statistical fluctuations play in creating avalanches – patterns of complex bursting activity with scale-free properties – in leaky Markovian networks, which are special cases of Markovian reaction networks. Using this class of models, we develop a probabilistic approach that employs a potential energy landscape perspective coupled with a macroscopic description based on statistical thermodynamics. We identify six important thermodynamic quantities essential for characterizing system behavior as a function of network size Ω: the internal potential energy, entropy, free potential energy, internal pressure, pressure, and bulk modulus. In agreement with classical phase transitions, these quantities evolve smoothly as a function

of Ω until a critical network size Ω_c is reached. At Ω_c, a discontinuity in pressure is observed that leads to a spike in the bulk modulus demarcating loss of thermodynamic robustness. We attribute this novel result to a reallocation of the ground states (global minima) of the system's stationary potential energy landscape to a noise-induced deformation of its topographic surface. Further analysis demonstrates that avalanching is a complex mode of operation that dominates system dynamics at near-critical or subcritical network sizes caused by appreciable levels of intrinsic noise. Illustrative examples are provided using an epidemiological model of bacteria infection, where avalanching has never been characterized, and a previously studied model of computational neuroscience, where avalanching was erroneously attributed to specific neural architectures. The general methods developed here can be used to study the emergence of bursting (and other complex phenomena) in many biological, physical and man-made interaction networks.

5.1 Motivation

An important problem in many scientific disciplines is understanding how extrinsic and intrinsic factors enable a complex physical system to exhibit a bursting behavior that leads to avalanching [167,168]. Avalanching is a form of spontaneous comportment characterized by irregular and isolated bursts of activity that follow a scale-free distribution typical to systems near criticality. In the brain, this mode of operation is thought to play a crucial role in information processing, storage, and learning [168, 169].

Although avalanche dynamics have been extensively studied *in vitro* and *in vivo* for cortical neural networks [168], it is not clear which are the underlying causes of this behavior. A recent *in silico* attempt to address this issue [19] was based on approximating the dynamics of a Markovian model of nonlinear interactions between noisy excitatory and inhibitory neurons by Gaussian fluctuations around the macroscopic (mean) system behavior using the linear noise approximation (LNA) method of van Kampen [1]. This led to the conclusion that the cause of neural avalanches

is a balanced feed-forward (BFF) network structure. We argue in this chapter that the Gaussian approximation used to arrive at this conclusion is not appropriate for studying avalanching, thus leading to deficient results. As a consequence, understanding the underlying causes of avalanching *in silico* is still an open problem.

To address this challenge, we introduce a theoretical framework that allows us to examine the role of intrinsic noise in inducing critical behavior that leads to avalanching. We focus on a special set of Markovian reaction models, which we term *leaky Markovian network* (LMN), with binary-valued (0,1) state dynamics. These dynamics are described by a time-dependent probability distribution that evolves according to a well-defined master equation [5]. It turns out that a LMN is a continuous-time stochastic Boolean network model with a state-dependent asynchronous node updating scheme. LMNs can model a number of natural and man-made systems of interacting species, such as gene networks, neural networks, epidemiological networks, and social networks.

Recent work has clearly demonstrated the importance of modeling physical systems stochastically using Markovian networks, since the intrinsic noise produced by these networks may induce behavior not accounted for by deterministic models [106, 170, 171]. Examples of such behavior include the emergence of noise-induced modes, stochastic transitions between different operational states, and "stabilization" of existing modes.

By using a LMN model, we study the effect of intrinsic noise on bursting. We do so by employing the notion of potential energy landscape [5, 96, 172] and by establishing a connection between statistical thermodynamics and the kinetics of bursting. We quantify the landscape by calculating logarithms of the ratios between the stationary probabilities of individual states and the stationary probability of the most probable state (details in Section 5.2). To reduce computational complexity, we follow a coarse graining approach that transforms the original LMN model into another (non-binary) LMN model with appreciably smaller state-space. To accomplish this task, we partition the nodes of the LMN into a number of homogeneous subpopulations and use the dynamic evolution of the fraction of the active nodes (nodes with value 1) in each subpopulation

to characterize system behavior. Moreover, we parameterize the LMN in terms of the network size $\Omega = N/N_0$, where N is the net number of nodes in the network and $N_0 \gg 1$ is a normalizing constant such that Ω can be approximately considered to be continuous-valued.

The behavior of the fractional activity process is fundamentally affected by Ω. In general, the strength of stochastic fluctuations (intrinsic noise) in the activity process may be thought of as the probability of moving uphill on a fixed potential energy surface, and this probability decays exponentially with increasing Ω. At sufficiently large network sizes Ω, the LMN operates around a ground state of the potential surface located at a fixed point $\boldsymbol{\mu}^*$ predicted by the macroscopic equations associated with the LNA method. However, as the network size decreases, a new mode of operation is introduced in the system in the form of a potential well in the topographic surface of the energy landscape, located at the *inactive* state $\mathbf{0}$. This is a "noise-induced" mode, since it appears at small network sizes at which the fractional activity process is subject to appreciable intrinsic fluctuations.

We show that noise-induced deformation of the stationary potential energy landscape is the underlying cause of bursting in LMNs. For sufficiently large network sizes, the potential energy landscape can be approximated by a quadratic surface centered at $\boldsymbol{\mu}^*$. In this case, the LMN operates within the potential well associated with this mode, except for rare and brief random excursions away from that mode. As a consequence, the fractional activity process will fluctuate in a Gaussian-like manner around the macroscopic mode. At smaller network sizes, the fractional activity process is characterized by a bistable behavior between the macroscopic and noise-induced modes, spending most time within the potential well associated with the macroscopic mode, at which the potential energy surface attains its global minimum, while occasionally jumping inside the potential well associated with the noise-induced mode at $\mathbf{0}$. As a consequence, the fractional activity dynamics take on a bursting behavior characterized by long periods of appreciable activity followed by short periods of minimal (almost zero) activity. When the network size decreases further, the noise-induced mode becomes the main stable operating point (i.e., the point at which the potential energy surface attains

its global minimum), whereas the macroscopic mode becomes shallower and eventually disappears. In this case, the system is trapped within the potential well associated with the noise-induced mode, except for random and brief excursions away from that mode. As a consequence, the fractional activity process will still exhibit bursting, but now characterized by long periods of minimal (almost zero) activity followed by short bursts of appreciable activity.

Thermodynamic analysis reveals critical behavior in LMNs (details in Section 5.2). By employing a number of statistical thermodynamic quantities, such as internal and free potential energies, entropy, internal pressure, pressure and bulk modulus (inverse compressibility), we effectively summarize the stochastic behavior of a LMN as its size Ω decreases to zero. We also use these summaries to quantify network robustness and the stability of a given state. In agreement with the classical theory of phase transitions, the previous thermodynamic quantities evolve smoothly as a function of Ω until a critical network size Ω_c is reached. At this size, a discontinuity is observed in the system pressure, which produces a spike in the bulk modulus demarcating loss of thermodynamic robustness. Critical behavior is caused by reallocation of the ground states (global minima) of the potential energy landscape due to noise-induced deformation of its topographic surface. In particular, observed critical behavior produces two distinct phases: one in which the fixed point $\boldsymbol{\mu}^*$ predicted by the macroscopic equations associated with the LNA method constitutes the ground state of the potential energy landscape and one in which the ground state is reallocated to the noise-induced mode at $\mathbf{0}$. We conclude that avalanching is a complex mode of operation that dominates system dynamics at near-critical and subcritical network sizes due to deformations of the potential energy landscape as the network size decreases to zero, caused by appreciable levels of intrinsic noise.

5.2 LMN theory and analysis

5.2.1 Leaky Markovian networks

We consider a directed weighted network $\mathcal{G}$ with N nodes from a set $\mathcal{N} = \{1, 2, \ldots, N\}$, characterized by an $N \times N$ adjacency matrix $\mathbb{A}$. The element $a_{nn'}$ of this matrix assigns a value to the edge leaving the n'-th node and entering the n-th node whose importance will become clear shortly. Each node represents a species (e.g., an individual or neuron) which, in some well-defined sense, can be active or inactive at time t with some probability. We use $X_n(t)$ to denote the state of the n-th node of the network at time t, taking value 1 if the node is active and 0 if the node is inactive. Then, we represent the state dynamics of the network by an N-dimensional random process $\{\boldsymbol{X}(t),\, t \geq 0\}$ whose n-th element $X_n(t)$ takes binary 0-1 values. We refer to $\{\boldsymbol{X}(t),\, t \geq 0\}$ as the activity process.

We assume that, within an infinitesimally small time interval $[t, t + dt)$, the state of the n-th node is influenced by the net input $r_n(\boldsymbol{x})$ to the node, where $\boldsymbol{x}$ is the state of the network $\mathcal{G}$ at time t and r_n is a real-valued scalar function. In particular, we assume that the probability of the n-th node to transition from the inactive to the active state within $[t, t + dt)$ is proportional to dt, given by $p_n^+(\boldsymbol{x})dt + o(dt)$, where $p_n^+(\boldsymbol{x})$ is known as the propensity function and $o(dt)$ is a term that goes to zero faster than dt. We set

$$p_n^+(\boldsymbol{x}) = (1 - x_n)[\ell_n^+ + f_n(r_n(\boldsymbol{x}))], \tag{5.2.1}$$

for some nonnegative parameter ℓ_n^+ and a nonnegative function $f_n(r)$. The term $(1 - x_n)$ ensures that transition to the active state is possible only when the n-th node is inactive (i.e., when $x_n = 0$), whereas the function $f_n(r)$ describes how the net input affects the probability of transition. On the other hand, when $\ell_n^+ > 0$, the parameter ℓ_n^+ forces the node to be "leaky," in the sense that it has a fixed propensity to transition from the inactive to the active state, even when the net input is zero. "Leakiness" is a property observed in many applications, including the ones discussed in this chapter.

We also assume that the probability of the n-th node to transition from the active to the inactive state within $[t, t+dt)$ is given by $p_n^-(\boldsymbol{x})dt + o(dt)$, where the propensity function $p_n^-(\boldsymbol{x})$ is given by

$$p_n^-(\boldsymbol{x}) = x_n[\ell_n^- + g_n(r_n(\boldsymbol{x}))] \tag{5.2.2}$$

for some nonnegative parameter ℓ_n^- and a nonnegative function $g_n(r)$. The term x_n ensures that transition to the inactive state is possible only when the n-th node is active (i.e., when $x_n = 1$), whereas the function $g_n(r)$ describes how the net input affects the probability of transition. On the other hand, when $\ell_n^- > 0$, the parameter ℓ_n^- forces the node to be "leaky," in the sense that it has a fixed propensity to transition from the active to the inactive state even when the net input is zero.

In general, the weights $a_{nn'}$ are used to determine the net input $r_n(\boldsymbol{x})$ to node n. As a matter of fact, $r_n(\boldsymbol{x})$ must not depend on $x_{n'}$ when $a_{nn'} = 0$. In particular, we set $a_{nn} = 0$, for every $n \in \mathcal{N}$, which implies that the nodes are not self-regulating. In some applications (such as the ones considered in this chapter), we can set

$$r_n(\boldsymbol{x}) = h_n + \boldsymbol{a}_n^T \boldsymbol{x}, \tag{5.2.3}$$

where $\boldsymbol{a}_n^T$ is the n-th row of the adjacency matrix $\mathbb{A}$ and h_n is a constant. In this case, h_n may represent the influence of external sources on the node (which we assume for simplicity to be fixed and known), whereas $\boldsymbol{a}_n^T \boldsymbol{x}$ represents the influence of all active nodes in the network on the state of the n-th node.

The process $\{\boldsymbol{X}(t),\ t \geq 0\}$ is Markovian. By assuming that all nodes in the network are initially inactive at time $t = 0$, we can show that the probability distribution $P(\boldsymbol{x}; t) := \Pr[\boldsymbol{X}(t) = \boldsymbol{x} \mid \boldsymbol{X}(0) = \boldsymbol{0}]$ satisfies the master equation

$$\frac{\partial P(\boldsymbol{x}; t)}{\partial t} = \sum_{n=1}^{N} \{p_n^+(\boldsymbol{x} - \boldsymbol{e}_n) P(\boldsymbol{x} - \boldsymbol{e}_n; t) + p_n^-(\boldsymbol{x} + \boldsymbol{e}_n) P(\boldsymbol{x} + \boldsymbol{e}_n; t) - [p_n^+(\boldsymbol{x}) + p_n^-(\boldsymbol{x})] P(\boldsymbol{x}; t)\}, \tag{5.2.4}$$

for $\boldsymbol{x} \in \mathcal{X} := \{0, 1\}^N$, initialized with the Kronecker delta function $\Delta(\boldsymbol{x})$ [i.e., $P(\boldsymbol{x}; 0) = \Delta(\boldsymbol{x})$], where $\boldsymbol{e}_n$ is the n-th column of the $N \times N$ identity matrix. The model described by this equation is a continuous Boolean network model with state-dependent asynchronous node updating (details in

the next section). Unfortunately, solving this equation is a notoriously difficult task, especially when the number N of nodes in the network is large. This is due to the fact that we need to calculate the probabilities $P(\boldsymbol{x};t)$, for $t>0$, at every point $\boldsymbol{x}$ in the state space $\mathcal{X}$, whose cardinality $|\mathcal{X}|$ grows exponentially as a function of N, since $|\mathcal{X}|=2^N$.

5.2.2 LMNs, Markovian reaction networks, and Boolean networks

LMNs are simply a special case of Markovian reaction networks. In particular, we note that the following reactions characterize every LMN,

$$\sum_{n'\neq n} a_{nn'}X_{n'} \rightleftarrows X_n + \sum_{n'\neq n} a_{nn'}X_{n'}, \tag{5.2.5}$$

for all $n\in\mathcal{N}$, with propensity functions $p_n^+(\boldsymbol{x})$ and $p_n^-(\boldsymbol{x})$, respectively for the forward and reverse reaction.

Boolean networks were introduced by S. Kauffman more than 40 years ago as models of gene regulation [173,174]. Since then, they have been extensively used as simple models for the dynamics of complex networks. A Boolean network is typically formulated as a directed graph with N nodes in a set $\mathcal{N}=\{1,2,\ldots,N\}$, whose state of the n-th node is characterized by a (deterministic) binary variable $x_n(t)$ taking 0-1 values. The status of each node is influenced by an input function $b_n(\boldsymbol{x})$, which is a Boolean function over a subset of the binary state variables $x_1,x_2,\ldots,x_N$. Given a fixed time step Δt, the state of the Boolean network at time $(m+1)\Delta t$ is determined by *synchronously* updating the states of all nodes in the network at time $m\Delta t$ using the deterministic rule

$$x_n((m+1)\Delta t) = b_n(\boldsymbol{x}(m\Delta t)), \quad \text{for every } n\in\mathcal{N}, \tag{5.2.6}$$

or more general probabilistic updating rules [175,176].

The vast majority of complex networks of interest do not update their states in a synchronous manner. As a consequence, Boolean networks tend to oversimplify the dynamics of many real networks. To address this problem, a number of investigators have focused their effort on

an *asynchronous* stochastic updating scheme that leads to stochastic asynchronous Boolean networks [177–179]. According to this scheme, the state of a Boolean network at time $(m+1)\Delta t$ is determined by randomly selecting a node in the network (usually uniformly among all nodes), by updating the state of this node using the associated Boolean function, and by leaving the states of the remaining nodes unchanged. In this case,

$$x_n((m+1)\Delta t) = \begin{cases} b_{n^*}(\boldsymbol{x}(m\Delta t)), & \text{for } n = n^* \\ \\ x_n(m\Delta t), & \text{for } n \neq n^*, \end{cases} \tag{5.2.7}$$

where n^* is the node selected to be updated at time $(m+1)\Delta t$.

Although the previous modification results in a model than may be more realistic than the classical Boolean model, it does not take into account major features of real complex networks. In particular, the model does not account for the facts that state updating can occur at any time t (not necessarily at discrete times $m\Delta t$) and that the time of next updating as well as the node to be updated can be influenced by the current state of the network. To address these issues, a number of models have been proposed in the literature [180–186]. It turns out that the LMN model discussed in this chapter effectively addresses these problems and provides a natural alternative to the stochastic Boolean network models studied in the literature; see [186]. As a matter of fact, if the network is at state $\boldsymbol{x}(t)$ at time t, then the time $t+\tau^*$ at which the state of the network will be next updated can be determined by drawing a sample τ^* from the exponential distribution

$$e_t(\tau) = \Big\{ \sum_{n\in\mathcal{N}} \alpha_n(\boldsymbol{x}(t)) \Big\} \exp\Big\{ -\tau \sum_{n\in\mathcal{N}} \alpha_n(\boldsymbol{x}(t)) \Big\},\ \tau > 0, \tag{5.2.8}$$

where

$$\alpha_n(\boldsymbol{x}) = (1-x_n)[\ell_n^+ + f_n(r_n(\boldsymbol{x}))] + x_n[\ell_n^- + g_n(r_n(\boldsymbol{x}))]. \tag{5.2.9}$$

Moreover, the node to be updated can be specified by drawing a sample n^* from the probability distribution

$$u_t(n) = \frac{\alpha_n(\boldsymbol{x}(t))}{\sum_{n'\in\mathcal{N}} \alpha_{n'}(\boldsymbol{x}(t))},\ n \in \mathcal{N}. \tag{5.2.10}$$

In this case [compare with Eq. (5.2.7)]

$$x_n(t+\tau^*) = \begin{cases} 1 - x_{n^*}(t), & \text{for } n = n^* \\ x_n(t), & \text{for } n \neq n^*. \end{cases} \tag{5.2.11}$$

This implies that the LMN model is a continuous-time Boolean network model with state-dependent asynchronous node updating and Boolean functions $b_n(\boldsymbol{x}) = 1 - x_n$ assigned at each node $n \in \mathcal{N}$.

5.2.3 Coarse graining

We can address the previously mentioned problem of exponential growth of the state space by employing a "coarse graining" procedure which allows us to appreciably reduce the size of the state space while retaining key properties of the system under consideration. We assume that we can partition the population $\mathcal{N} = \{1, 2, \ldots, N\}$ of all species in the network $\mathcal{G}$ into K homogenous sub-populations $\mathcal{N}_k$, $k = 1, 2, \ldots, K$, where $K \ll N$. Due to the homogeneity of each sub-population, it may not be of particular interest to track the states of individual species in a given sub-population $\mathcal{N}_k$. Instead, it may be sufficient to track the fraction $Y_k(t)$ of active species in $\mathcal{N}_k$, defined by

$$Y_k(t) \;:=\; \frac{1}{N_k} \sum_{n \in \mathcal{N}_k} X_n(t), \tag{5.2.12}$$

where $N_k := |\mathcal{N}_k|$. In this case, we may replace the original network with a smaller directed weighted network $\mathcal{G}_0$ comprised of K nodes from the set $\mathcal{K} = \{1, 2, \ldots, K\}$ that represent the homogeneous sub-populations. We assume that, for every $k \in \mathcal{K}$, there exists a function ρ_k such that $r_n(\boldsymbol{x}) = \rho_k(\boldsymbol{y})$, for all $n \in \mathcal{N}_k$, where $\boldsymbol{y}$ is a $K \times 1$ vector whose k-th element y_k is given by $y_k = N_k^{-1} \sum_{n \in \mathcal{N}_k} x_n$. In this case, the stochastic process $\{\boldsymbol{Y}(t),\, t \geq 0\}$ is also Markovian, governed by the following master equation (details in Appendix D)

$$\frac{\partial P(\boldsymbol{y};t)}{\partial t} \;=\; \sum_{k \in \mathcal{K}} \{\pi_k^+(\boldsymbol{y} - \widetilde{\boldsymbol{e}}_k) P(\boldsymbol{y} - \widetilde{\boldsymbol{e}}_k; t) + \pi_k^-(\boldsymbol{y} + \widetilde{\boldsymbol{e}}_k) P(\boldsymbol{y} + \widetilde{\boldsymbol{e}}_k; t) - [\pi_k^+(\boldsymbol{y}) + \pi_k^-(\boldsymbol{y})] P(\boldsymbol{y}; t)\}, \tag{5.2.13}$$

initialized with the Kronecker delta function $\Delta(\boldsymbol{y})$ [i.e., $P(\boldsymbol{y}; 0) = \Delta(\boldsymbol{y})$], where $P(\boldsymbol{y}; t) := \Pr[\boldsymbol{Y}(t) = \boldsymbol{y} \mid \boldsymbol{Y}(0) = \boldsymbol{0}]$ and $\widetilde{\boldsymbol{e}}_k$ is the k-th column of the $K \times K$ identity matrix multiplied by N_k^{-1}. The new

propensity functions are given by

$$\pi_k^+(\boldsymbol{y}) = N_k(1-y_k)[\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{y}))] \tag{5.2.14}$$

$$\pi_k^-(\boldsymbol{y}) = N_k y_k[\lambda_k^- + \gamma_k(\rho_k(\boldsymbol{y}))], \tag{5.2.15}$$

where ϕ_k, γ_k, λ_k^+, and λ_k^- are such that, for every $k \in \mathcal{K}$, $f_n = \phi_k, g_n = \gamma_k, \ell_n^+ = \lambda_k^+$, and $\ell_n^- = \lambda_k^-$, for $n \in \mathcal{N}_k$. We refer to $\{\boldsymbol{Y}(t),\, t \geq 0\}$ as the fractional activity process.

When the input to a node n of the network is given by $r_n(\boldsymbol{x}) = h_n + \boldsymbol{a}_n^T\boldsymbol{x}$, there is indeed a function $\rho_k(\boldsymbol{y})$ so that $r_n(\boldsymbol{x}) = \rho_k(\boldsymbol{y})$, for every $n \in \mathcal{N}_k$. This function is given by $\rho_k(\boldsymbol{y}) = \eta_k + \sum_{k' \in \mathcal{K}} w_{kk'} y_{k'}$, where η_k and $w_{kk'}$ are such that $h_n = \eta_k$, for every $n \in \mathcal{N}_k$, and $a_{nn'} = w_{kk'}/N_{k'}$, for every $n \in \mathcal{N}_k$, $n' \in \mathcal{N}_{k'}$ (details in Appendix D). Note also that $Y_k(t)$ takes values in $\mathcal{Y}_k := \{0, 1/N_k, \ldots, 1\}$. Therefore, the fractional activity process $\{\boldsymbol{Y}(t),\, t \geq 0\}$ takes values in $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_K$. As a result, the state-space $\mathcal{Y}$ will be appreciably smaller than $\mathcal{X}$, since $|\mathcal{Y}| = \prod_{k=1}^K (1+N_k) \ll 2^N = |\mathcal{X}|$, and solving the master equation of the fractional activity process will be easier than solving the master equation of the activity process.

5.2.4 Macroscopic equations and LNA

We define the fractional "size" of the k-th sub-population as $\zeta_k := N_k/N$. The thermodynamic limit is obtained by taking $N_k \to \infty$, for every $k \in \mathcal{K}$, such that all ζ_k's remain fixed. In this case, $\boldsymbol{Y}(t)$ becomes a continuous random variable in the K-dimensional closed unit hypercube $[0,1]^K$. Furthermore, since $N_k \to \infty$, for every $k \in \mathcal{K}$, one might expect that the intrinsic noise at each node of the coarse network $\mathcal{G}_0$ will be averaged out due to coarse graining. As a matter of fact, it can be shown that $\boldsymbol{Y}(t)$ converges in distribution to the deterministic solution $\boldsymbol{\mu}(t)$ of the macroscopic differential equations

$$\frac{d\mu_k(t)}{dt} = [1-\mu_k(t)][\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{\mu}(t)))] - \mu_k(t)[\lambda_k^- + \gamma_k(\rho_k(\boldsymbol{\mu}(t)))], \tag{5.2.16}$$

$t > 0, k \in \mathcal{K}$, initialized by $\boldsymbol{\mu}(0) = \mathbf{0}$. For simplicity of notation, we denote the thermodynamic limit by $N \to \infty$.

If the macroscopic equations have a unique and stable fixed point $\boldsymbol{\mu}^*$ in the interior of the unit hypercube $[0,1]^K$, then for large enough but finite N, the LNA method allows us to approximate the fractional activity process $\boldsymbol{Y}(t)$ by adding correlated Gaussian noise $\boldsymbol{W}(t)$ to the macroscopic solution $\boldsymbol{\mu}(t)$. In this case,

$$Y_k(t) \simeq \mu_k(t) + \frac{W_k(t)}{\sqrt{N_k}}, \quad t > 0, k \in \mathcal{K}, \tag{5.2.17}$$

where, for each $t, W_k(t), k \in \mathcal{K}$, are zero-mean correlated Gaussian random variables with correlations $r_{kk'}(t) = \mathrm{E}[W_k(t)W_{k'}(t)]$ that satisfy a system of Lyapunov equations (details in Appendix D). As a consequence, $\boldsymbol{Y}(t)$ is approximated by a multivariate Gaussian random vector with mean $\boldsymbol{\mu}(t)$ and covariance matrix $\mathbb{C}(t) = \mathbb{N}\mathbb{R}(t)\mathbb{N}^T$, where $\mathbb{N}$ is a diagonal matrix with elements $1/\sqrt{N_1}, 1/\sqrt{N_2}, \ldots, 1/\sqrt{N_K}$, and $\mathbb{R}(t)$ is the correlation matrix of random vector $\boldsymbol{W}(t)$.

5.2.5 Thermodynamic stability, robustness, and critical behavior

We consider the probability distribution $P_\Omega(\boldsymbol{y};t)$ of the fractional activity process $\boldsymbol{Y}(t)$ at time t, where we explicitly denote the dependence of this distribution on the network size Ω. Let $\boldsymbol{y}^*_\Omega(t)$ be a state in $\mathcal{Y}$ at which $P_\Omega(\boldsymbol{y};t)$ attains its (global) maximum value at time t and define the function

$$V_\Omega(\boldsymbol{y};t) := -\frac{1}{\Omega}\ln\frac{P_\Omega(\boldsymbol{y};t)}{P_\Omega(\boldsymbol{y}^*_\Omega(t);t)} \tag{5.2.18}$$

Note that $V_\Omega(\boldsymbol{y};t) \geq 0$ with equality if and only if $\boldsymbol{y}$ is a state at which $P_\Omega(\boldsymbol{y};t)$ attains its (global) maximum value, known as a ground state. Moreover,

$$P_\Omega(\boldsymbol{y};t) = \frac{1}{Z_\Omega(t)}\, e^{-\Omega V_\Omega(\boldsymbol{y};t)}, \tag{5.2.19}$$

for $\boldsymbol{y} \in \mathcal{Y}$, $t \geq 0$, where

$$Z_\Omega(t) := \sum_{\boldsymbol{y}\in\mathcal{Y}} e^{-\Omega V_\Omega(\boldsymbol{y};t)} \tag{5.2.20}$$

In this case, $P_\Omega(\boldsymbol{y};t)$ is a Boltzmann-Gibbs distribution with potential energy function $V_\Omega(\boldsymbol{y};t)$ and partition function $Z_\Omega(t)$. The (local or global) minima of $V_\Omega(\boldsymbol{y};t)$ are associated with "potential

wells" (basins of attraction) in the energy surface, which correspond to peaks in the probability distribution $P_\Omega(\boldsymbol{y};t)$. We may therefore view the fractional activity dynamics as fluctuations on a time-evolving potential energy landscape $V_\Omega(\boldsymbol{y};t)$ in the multidimensional state-space $\mathcal{Y}$, where downhill motions (towards the bottom of a potential well) are preferred with high probability, but random uphill motions can also occur with increasing probability as the population size decreases.

We are interested in the stationary potential energy $\overline{V}_\Omega(\boldsymbol{y}) := \lim_{t\to\infty} V_\Omega(\boldsymbol{y};t)$, since its landscape remains fixed once the stochastic dynamics reach this point. At steady-state, the fractional activity dynamics simply perform a random walk on $\overline{V}_\Omega(\boldsymbol{y})$. To compute $V_\Omega(\boldsymbol{y};t)$ and $\overline{V}_\Omega(\boldsymbol{y})$, we solve the master equation of the fractional activity process numerically.

Based on the partition function, we can define a number of quantities that can be used to characterize the behavior of the LMN [and, as a matter of fact, of any Markovian network [5]], when nodes are removed from the network. We adopt these quantities from classical thermodynamics where they are used to describe the behavior of a physical system as its volume contracts or expands [15,115,187,188]. In particular, we can define the internal energy, entropy, and Helmholtz free energy, and subsequently introduce the concepts of internal potential energy, free potential energy, internal pressure, pressure, and bulk modulus. In the following, we denote by $\overline{A}$ the stationary limit of a time-varying parameter $A(t)$; i.e., $\overline{A} = \lim_{t\to\infty} A(t)$.

The internal energy at time t is defined by

$$\mathcal{U}_\Omega(t) := \Omega \mathrm{E}_t[U_\Omega(\boldsymbol{Y})], \quad \text{for} \quad t \geq 0, \tag{5.2.21}$$

where

$$U_\Omega(\boldsymbol{y}) := -\frac{1}{\Omega} \ln \overline{P}_\Omega(\boldsymbol{y}) \tag{5.2.22}$$

is the energy of state $\boldsymbol{y} \in \mathcal{Y}$ and $\mathrm{E}_t[\cdot]$ denotes expectation with respect to the probability distribution $P_\Omega(\boldsymbol{y};t)$. On the other hand, the Helmholtz free energy $\mathcal{F}_\Omega(t)$ at time t is defined by

$$\mathcal{F}_\Omega(t) := \mathcal{U}_\Omega(t) - \mathcal{S}_\Omega(t), \tag{5.2.23}$$

where $\mathcal{S}_\Omega(t)$ is the entropy at time t, given by

$$\mathcal{S}_\Omega(t) := -\mathrm{E}_t[\ln P_\Omega(\boldsymbol{Y};t)]. \tag{5.2.24}$$

In thermodynamic terms, the Helmholtz free energy measures the energy available in the LMN to do work when the number of nodes is kept fixed. Note that $\mathcal{U}_\Omega(t) \geq 0$ and $0 \leq \mathcal{S}_\Omega(t) \leq \ln|\mathcal{Y}|$, for every Ω and $t \geq 0$, where $|\mathcal{Y}|$ is the cardinality of the state-space $\mathcal{Y}$. It also turns out that $\mathcal{F}_\Omega(t) \geq 0$ and $d\mathcal{F}_\Omega(t)/dt \leq 0$, for every Ω and $t \geq 0$, with equality only at steady-state [5, 117, 126].

From Eq. (5.2.18), Eq. (5.2.21), and Eq. (5.2.22), we can show that

$$\mathcal{U}_\Omega(t) = \mathcal{V}_\Omega(t) + \Omega U_\Omega(\overline{\boldsymbol{y}}^*_\Omega), \tag{5.2.25}$$

where

$$\mathcal{V}_\Omega(t) := \Omega \mathrm{E}_t[\overline{V}_\Omega(\boldsymbol{Y})], \quad \text{for} \;\; t \geq 0, \tag{5.2.26}$$

is the internal potential energy of the LMN. Moreover,

$$\mathcal{F}_\Omega(t) := \mathcal{A}_\Omega(t) + \Omega U_\Omega(\overline{\boldsymbol{y}}^*_\Omega), \tag{5.2.27}$$

by virtue of Eq. (5.2.23) and Eq. (5.2.25), where

$$\mathcal{A}_\Omega(t) := \mathcal{V}_\Omega(t) - \mathcal{S}_\Omega(t). \tag{5.2.28}$$

Note that $\mathcal{A}_\Omega(t)$ is the portion of the Helmholtz free energy due to the internal potential energy of the LMN. For this reason, we refer to $\mathcal{A}_\Omega(t)$ as the free potential energy. In thermodynamic terms, the free potential energy measures the portion of the energy, not accounted for by the energy of the most likely state, available in the LMN to do work when the number of nodes is kept fixed.

The quantities

$$\mathcal{P}^0_\Omega(t) := \frac{\partial \mathcal{V}_\Omega(t)}{\partial \Omega} \tag{5.2.29}$$

$$\mathcal{P}_\Omega(t) := -\frac{\partial \mathcal{A}_\Omega(t)}{\partial \Omega} \tag{5.2.30}$$

$$\mathcal{B}_\Omega(t) := -\Omega \frac{\partial \mathcal{P}_\Omega(t)}{\partial \Omega} \tag{5.2.31}$$

define the internal pressure, pressure, and bulk modulus of the LMN, repsctively. The internal pressure quantifies the rate of change in internal potential energy with respect to a change in the number of nodes, whereas the pressure quantifies the rate of change in free potential energy. Moreover, the bulk modulus measures the network's resistance to changing pressure. Note that $\mathcal{P}_\Omega(t) = \mathcal{Q}_\Omega(t) - \mathcal{P}^0_\Omega(t)$, where

$$\mathcal{Q}_\Omega(t) := \frac{\partial \mathcal{S}_\Omega(t)}{\partial \Omega} \tag{5.2.32}$$

is the rate of entropy change with respect to a change in the number of nodes. Moreover, a network with near zero bulk modulus experiences negligible changes in pressure under changes in the number of nodes. The inverse bulk modulus is known as *compressibility*.

In the following, we focus our interest on the stationary behavior of a LMN. Due to the irreducibility properties of LMNs, all stationary thermodynamic quantities are unique and characteristic to the particular network under consideration. From Eq. (5.2.18) and Eq. (5.2.26), note that

$$\begin{aligned}\overline{\mathcal{V}}_\Omega &= \mathrm{E}[-\ln \overline{P}_\Omega(\boldsymbol{Y})] - [-\ln \overline{P}_\Omega(\overline{\boldsymbol{y}}^*_N)] \\ &= \mathrm{E}[\mathcal{I}_\Omega(\boldsymbol{Y})] - \mathcal{I}_\Omega(\overline{\boldsymbol{y}}^*_\Omega) \geq 0,\end{aligned} \tag{5.2.33}$$

where $\mathrm{E}[\cdot]$ denotes expectation with respect to the stationary probability distribution $\overline{P}_\Omega(\boldsymbol{y})$ and

$$\mathcal{I}_\Omega(\boldsymbol{y}) := -\ln \overline{P}_\Omega(\boldsymbol{y}) = \Omega U_\Omega(\boldsymbol{y}). \tag{5.2.34}$$

$\mathcal{I}_\Omega(\boldsymbol{y})$ quantifies the amount of information associated with the occurrence of state $\boldsymbol{y}$ at steady-state, known as the self-information of state $\boldsymbol{y}$. Therefore, and from an information-theoretic perspective, the internal potential energy measures how far the self-information of the most likely state at steady-state is from the expected self-information of all network states (which is the entropy). Note that zero internal potential energy implies zero self-information for the most likely state. In this case, the network will be at the most likely state with probability one. As a consequence, we may consider the internal potential energy as a thermodynamic measure of the "stability" of a particular ground

state of the stationary potential energy landscape $\overline{V}_{\Omega}(\boldsymbol{y})$ with smaller values indicating increasing stability of that state.

From Eq. (5.2.19), Eq. (5.2.24), Eq. (5.2.26), Eq. (5.2.28), and Eq. (5.2.34), we also have that

$$\begin{aligned}
\overline{\mathcal{A}}_{\Omega} &= \sum_{\boldsymbol{y}\in\mathcal{Y}} \Omega \overline{V}_{\Omega}(\boldsymbol{y}) \overline{P}_{\Omega}(\boldsymbol{y}) + \sum_{\boldsymbol{y}\in\mathcal{Y}} \big[\ln \overline{P}_{\Omega}(\boldsymbol{y})\big] \overline{P}_{\Omega}(\boldsymbol{y}) \\
&= \sum_{\boldsymbol{y}\in\mathcal{Y}} \Omega \overline{V}_{\Omega}(\boldsymbol{y}) \overline{P}_{\Omega}(\boldsymbol{y}) - \sum_{\boldsymbol{y}\in\mathcal{Y}} \Omega \overline{V}_{\Omega}(\boldsymbol{y}) \overline{P}_{\Omega}(\boldsymbol{y}) - \ln \overline{Z}_{\Omega} \\
&= -\ln \overline{Z}_{\Omega} \\
&= \ln \overline{P}_{\Omega}(\overline{\boldsymbol{y}}^{*}_{\Omega}) \\
&= -\mathcal{I}_{\Omega}(\overline{\boldsymbol{y}}^{*}_{\Omega}) \leq 0.
\end{aligned} \tag{5.2.35}$$

Therefore, the negative of the free potential energy is the self-information of the most likely state at steady-state. Note that the internal potential energy of an LMN with equally probable states at steady-state is zero, whereas its free potential energy equals $-\ln|\mathcal{Y}|$. On the other hand, the internal potential energy of an LMN with "crystalized" behavior around a *unique* ground state is also zero, and the same is true for the free potential energy. Note also that

$$\overline{\mathcal{P}}_{\Omega} = \frac{\partial \mathcal{I}_{\Omega}(\overline{\boldsymbol{y}}^{*}_{\Omega})}{\partial \Omega}, \tag{5.2.36}$$

by virtue of Eq. (5.2.30) and Eq. (5.2.35), whereas

$$\overline{\mathcal{B}}_{\Omega} = -\Omega \, \frac{\partial^2 \mathcal{I}_{\Omega}(\overline{\boldsymbol{y}}^{*}_{\Omega})}{\partial \Omega^2}, \tag{5.2.37}$$

by virtue of Eq. (5.2.31) and Eq. (5.2.36). Therefore, the pressure gives the slope of the support curve

$$\sigma_{*}(\Omega) := \mathcal{I}_{\Omega}(\overline{\boldsymbol{y}}^{*}_{\Omega}), \quad \Omega > 0, \tag{5.2.38}$$

of the self-information of the most likely state at steady-state, whereas the bulk modulus is proportional to the "curvature" of this curve.

In general,

$$\lim_{\Omega\to\infty} \overline{\mathcal{V}}_\Omega = v_\infty \,, \tag{5.2.39}$$

where v_∞ is a constant, given by

$$v_\infty = -\lim_{\Omega\to\infty} \mathrm{E}\left[\ln \frac{\overline{P}_\Omega(\boldsymbol{Y})}{\overline{P}_\Omega(\overline{\boldsymbol{y}}^*_\Omega)}\right], \tag{5.2.40}$$

whereas

$$\lim_{\Omega\to\infty} \overline{\mathcal{P}}^0_\Omega = \lim_{\Omega\to\infty} \overline{\mathcal{P}}_\Omega = \lim_{\Omega\to\infty} \overline{\mathcal{B}}_\Omega = 0 \,. \tag{5.2.41}$$

Indeed, Eq. (5.2.39) and Eq. (5.2.40) are a direct consequence of Eq. (5.2.18) and Eq. (5.2.26). On the other hand, Eq. (5.2.39) implies that $\lim_{\Omega\to\infty} \overline{\mathcal{P}}^0_\Omega = 0$ by virtue of Eq. (5.2.29). Note also that

$$\overline{\mathcal{P}}_\Omega = \frac{\partial \ln \overline{Z}_\Omega}{\partial \Omega} = \frac{\partial}{\partial \Omega} \ln\Big[\sum_{\boldsymbol{y}\in\mathcal{Y}} \exp\{-\Omega \overline{V}_\Omega(\boldsymbol{y})\}\Big], \tag{5.2.42}$$

by virtue of Eq. (5.2.20), Eq. (5.2.35), and Eq. (5.2.36). This implies

$$\lim_{\Omega\to\infty} \overline{\mathcal{P}}_\Omega = \frac{\partial \ln(1)}{\partial \Omega} = 0, \tag{5.2.43}$$

where we have used the fact that

$$\lim_{\Omega\to\infty} \exp\{-\Omega \overline{V}_\Omega(\boldsymbol{y})\} = \begin{cases} 1, & \text{if } \boldsymbol{y} = \overline{\boldsymbol{y}}^*_\Omega \\ \\ 0, & \text{otherwise} \,. \end{cases} \tag{5.2.44}$$

Clearly, the fact that $\lim_{\Omega\to\infty} \overline{\mathcal{P}}_\Omega = 0$ implies that $\lim_{\Omega\to\infty} \overline{\mathcal{B}}_\Omega = 0$, by virtue of Eq. (5.2.37).

The function

$$V_\infty(\boldsymbol{y}; t) := \lim_{\Omega\to\infty} V_\Omega(\boldsymbol{y}; t) \tag{5.2.45}$$

is known as large deviation rate function [189] and characterizes, as $\Omega \to \infty$, the decay rate of the probability distribution $P_\Omega(\boldsymbol{y}; t)$ away from the ground states. Moreover, its steady-state value $\overline{V}_\infty(\boldsymbol{y}) = \lim_{t\to\infty} V_\infty(\boldsymbol{y}; t)$ acts as a Lyapunov function for the macroscopic equations (5.2.16) [5,98]. This means that the solution $\boldsymbol{\mu}(t)$ of the macroscopic equations produces a downhill motion in the value of the potential energy landscape $\overline{V}_\infty(\boldsymbol{y})$ until it asymptotically reaches a stable stationary state.

Analytical derivation of $V_\infty(\boldsymbol{y};t)$ is not possible in general. However, when the macroscopic equations are *monostable*, the system size expansion of van Kampen implies that

$$\begin{aligned} V_\infty(\boldsymbol{y};t) &= \lim_{\Omega\to\infty} \frac{1}{2\Omega}\big(\boldsymbol{y}-\boldsymbol{\mu}(t)\big)^T \mathbb{C}^{-1}(t)\big(\boldsymbol{y}-\boldsymbol{\mu}(t)\big) \\ &= \lim_{\Omega\to\infty} \frac{1}{2\Omega}\big(\boldsymbol{y}-\boldsymbol{\mu}(t)\big)^T [\mathbb{N}\mathbb{R}(t)\mathbb{N}]^{-1}\big(\boldsymbol{y}-\boldsymbol{\mu}(t)\big) \\ &= \lim_{N\to\infty} \frac{N_0}{2}\big(\boldsymbol{y}-\boldsymbol{\mu}(t)\big)^T \mathbb{Z}\mathbb{R}^{-1}(t)\mathbb{Z}\big(\boldsymbol{y}-\boldsymbol{\mu}(t)\big) \\ &= \frac{N_0}{2}(\boldsymbol{y}-\boldsymbol{\mu}(t))^T \mathbb{Z}\mathbb{R}^{-1}(t)\mathbb{Z}(\boldsymbol{y}-\boldsymbol{\mu}(t)). \end{aligned} \tag{5.2.46}$$

In these equations, $\boldsymbol{\mu}(t)$ solves the macroscopic equations (5.2.16), the covariance matrix $\mathbb{C}(t)$ equals $\mathbb{N}\mathbb{R}(t)\mathbb{N}$, where $\mathbb{N}$ is a diagonal matrix with elements $1/\sqrt{N_1}, 1/\sqrt{N_2}, \ldots, 1/\sqrt{N_K}$ and $\mathbb{R}(t)$ is the correlation matrix of the random vector $\boldsymbol{W}(t)$ whose elements solve the Lyapunov equations

$$\frac{dr_{kk'}(t)}{dt} = D_k(\boldsymbol{\mu}(t))\Delta(k-k') + \sum_{k''\in\mathcal{K}} \sqrt{\frac{\zeta_k}{\zeta_{k''}}}\, A_{kk''}(\boldsymbol{\mu}(t))\, r_{k''k'}(t) + \sum_{k''\in\mathcal{K}} \sqrt{\frac{\zeta_{k'}}{\zeta_{k''}}}\, A_{k'k''}(\boldsymbol{\mu}(t))\, r_{kk''}(t), \tag{5.2.47}$$

for $t>0$ and $k,k'\in\mathcal{K}$, initialized with $r_{kk'}(0)=0$, for every $k,k'\in\mathcal{K}$, where Δ is the Kronecker delta function (see Appendix D), and $\mathbb{Z}$ is a diagonal matrix with elements $\zeta_1,\zeta_2,\ldots,\zeta_K$, where $\zeta_k = N_k/N$. Clearly, $V_\infty(\boldsymbol{y};t)$ is hyper-quadratic in this case around the macroscopic solution $\boldsymbol{\mu}(t)$, which is now the unique ground state. Moreover, the shapes of the equipotential surfaces centered at $\boldsymbol{\mu}(t)$ are ellipsoidal, determined by $\mathbb{R}^{-1}(t)$ and, therefore, by the Lyapunov equations (5.2.47).

If the stationary solution of the master equation can be well-approximated by the LNA method, then

$$\overline{\mathcal{V}}_\Omega \simeq \frac{K}{2}\,, \tag{5.2.48}$$

for those values of Ω at which this is true, where K is the dimension of the state-space $\mathcal{Y}$. Indeed,

from Eq. (5.2.26) and Eq. (5.2.46), we have that

$$\begin{aligned}
\overline{\mathcal{V}}_{\Omega} &\simeq \frac{1}{2}\mathrm{E}\big[(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})^{T}\overline{C}^{-1}(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})\big] \\
&= \frac{1}{2}\,\mathrm{tr}\Big[\mathrm{E}\big[(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})^{T}\overline{C}^{-1}(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})\big]\Big] \\
&= -\frac{1}{2}\mathrm{E}\Big[\mathrm{tr}\big[(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})^{T}\overline{C}^{-1}(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})\big]\Big] \\
&= \frac{1}{2}\mathrm{E}\Big[\mathrm{tr}\big[\overline{C}^{-1}(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})^{T}\big]\Big] \\
&= \frac{1}{2}\mathrm{tr}\Big[\overline{C}^{-1}\mathrm{E}\big[(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})^{T}\big]\Big] \\
&\simeq \frac{1}{2}\,\mathrm{tr}\big[\overline{C}^{-1}\overline{C}\big] = \frac{K}{2}\,,
\end{aligned} \tag{5.2.49}$$

where $\mathrm{tr}[A]$ denotes the trace of a matrix A. To obtain the previous result, we used three well-known properties of the trace: the trace of a scalar is itself, the trace is a linear operator (and therefore it commutes with expectation), and the trace is invariant under cyclic permutations. Moreover, we used the fact that $\mathrm{E}\big[(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})(\boldsymbol{Y}-\overline{\boldsymbol{\mu}})^{T}\big] \simeq \overline{C}$ when the LNA method provides a sufficiently accurate solution to the master equation at steady-state.

We can use the pressure as a measure of (thermodynamic) robustness of the LMN with respect to the network size Ω. We say that the LMN is robust against variations in network size if there is no appreciable change in pressure when adding or removing nodes. Therefore, the LMN is robust if the derivative $\partial\overline{\mathcal{P}}_{\Omega}/\partial\Omega$ of the pressure is small. As a consequence of Eq. (5.2.37), the LMN is robust if the bulk modulus is small (especially at small network sizes). This implies that a robust LMN must significantly resist changes in pressure. On the other hand, Eq. (5.2.37) and Eq. (5.2.38) reveal that the LMN is robust if the network is characterized by a "blunt" self-information curve $\sigma_{*}(\Omega)$ with small curvature. Note that, if Ω is sufficiently large, then Eq. (5.2.41) implies that the pressure and bulk modulus will approximately be zero and the network will be robust to changes in size. This also implies that the slope of the self-information support curve $\sigma_{*}(\Omega)$ will approximately be zero and the same will be true for its curvature.

It is important to emphasize here that we can use the bulk modulus $\overline{\mathcal{B}}_{\Omega}$ to detect network sizes at which the LMN exhibits critical behavior. As a matter of fact, it is well-known that an

intensive thermodynamic quantity, such as the pressure, may experience a sharp discontinuity when another thermodynamic variable, such as the network size, varies past a critical value. If the pressure $\overline{\mathcal{P}}_\Omega$ of the LMN experiences such a discontinuity as the network size Ω varies past a critical value Ω_c, then $\overline{\mathcal{B}}_\Omega$ will effectively capture this discontinuity by a delta function located at Ω_c, thus indicating that the network experiences phase transition at Ω_c.

As a consequence of the previous discussion, if the LNA method is valid for large values of Ω, then $\overline{\mathcal{B}}_\Omega \simeq 0$ at these network sizes. Moreover, if the thermodynamic behavior of the LMN changes abruptly when Ω is decreased past a critical value Ω_c, then $\overline{\mathcal{B}}_\Omega$ will produce a sharp spike at this value. A critical network size can demarcate a discontinuous transition of potential wells associated with the ground states. This is a direct consequence of the fact that spike-like behavior in the bulk modulus indicates an abrupt change in the slope of the self-information support curve $\sigma_*(\Omega)$ of the most likely state at steady-state, as predicted by Eq. (5.2.37).

5.2.6 Noise-induced modes, stochastic transitions and bursting

To explain why a noise-induced mode may appear at the origin of the state-space $\mathcal{Y}$ at low population sizes, let $T_e(\boldsymbol{y})$ be the *mean escape time* from a state $\boldsymbol{y} \in \mathcal{Y}$, defined as the average time required for the LMN to move from state $\boldsymbol{y}$ to any other state in $\mathcal{Y}$. Since the fractional activity process is Markovian, governed by the master equation (5.2.13), the time it spends at state $\boldsymbol{y}$ is an exponential random variable with rate parameter $\sum_{k\in\mathcal{K}}[\pi_k^+(\boldsymbol{y}) + \pi_k^-(\boldsymbol{y})]$, which implies that

$$T_e(\boldsymbol{y}) = \frac{1}{\sum_{k\in\mathcal{K}}\left[\pi_k^+(\boldsymbol{y}) + \pi_k^-(\boldsymbol{y})\right]} \ . \tag{5.2.50}$$

Clearly, if $T_e(\boldsymbol{y}) = \infty$, then the state $\boldsymbol{y}$ is *absorbing*. This means that, once the network reaches $\boldsymbol{y}$, it can never leave that state. As a consequence, we can use $T_e(\boldsymbol{y}^*)$ to assess the "stability" of a ground state $\boldsymbol{y}^*$ of the potential energy landscape of the LMN, with higher values of $T_e(\boldsymbol{y}^*)$ indicating that the state $\boldsymbol{y}^*$ is more stable. From Eq. (5.2.50) as well as Eq. (D.4) and Eq. (D.5) in Appendix D,

we have that

$$\begin{aligned}[T_e(\boldsymbol{y})]^{-1} &= \sum_{k\in\mathcal{K}} \big[N_k(1-y_k)\lambda_k^+ + N_k(1-y_k)\phi_k(\rho_k(\boldsymbol{y})) \\ &\qquad +N_k y_k\lambda_k^- + N_k y_k\gamma_k(\rho_k(\boldsymbol{y}))\big] \\ &= N\sum_{k\in\mathcal{K}} \big[\zeta_k(1-y_k)\lambda_k^+ + \zeta_k(1-y_k)\phi_k(\rho_k(\boldsymbol{y})) \\ &\qquad + \zeta_k y_k\lambda_k^- + \zeta_k y_k\gamma_k(\rho_k(\boldsymbol{y}))\big] \\ &= N\sum_{k\in\mathcal{K}} \zeta_k(1-y_k)\lambda_k^+ + \zeta_k y_k\lambda_k^- \geq 0. \end{aligned} \tag{5.2.51}$$

As a consequence, when $\lambda_k^+ = 0$ and $\lambda_k^- > 0$, for every $k \in \mathcal{K}$,[1] then $T_e(\boldsymbol{y}) = \infty$ only when $\boldsymbol{y} = \mathbf{0}$. If $T_e(\mathbf{0}) = \infty$, then the master equation (5.2.13) will have a trivial solution $P_\Omega(\boldsymbol{y};t) = \Delta(\boldsymbol{y})$, since the network is initialized at $\mathbf{0}$ and it will never move to another state. In this case, the resulting potential energy landscape $V_\Omega(\boldsymbol{y};t)$ will have a unique global minimum at $\mathbf{0}$ [as a matter of fact, $V_\Omega(\mathbf{0};t) = 0$, whereas $V_\Omega(\boldsymbol{y};t) = \infty$, for every $\boldsymbol{y} \neq \mathbf{0}$].

Since real-world networks must be characterized by non-trivial dynamics, we must have $T_e(\mathbf{0}) < \infty$. The fact that, when $T_e(\mathbf{0}) = \infty$, the probability distribution $P_\Omega(\boldsymbol{y};t)$ is concentrated at the origin suggests that a very large but finite $T_e(\mathbf{0})$ may be indicative of a probability distribution that assigns high probability to the zero state, creating a noise-induced mode at the origin of $\mathcal{Y}$. As a matter of fact, we expect that the stationary probability distribution $\overline{P}_\Omega(\boldsymbol{y})$ to have a peak around $\boldsymbol{y} = \mathbf{0}$ whenever the mean escape time of the fractional activity dynamics from the zero state is sufficiently large and the average time it takes for these dynamics to return to the zero state is small [see pages 100 & 101 in [190]].

In general, Eq. (5.2.51) implies that

$$[T_e(\mathbf{0})]^{-1} = N\sum_{k\in\mathcal{K}} \zeta_k\big[\lambda_k^+ + \phi_k(\rho_k(\mathbf{0}))\big], \tag{5.2.52}$$

where $\zeta_k = N_k/N$. Clearly, reducing the net population size N increases $T_e(\mathbf{0})$. It moreover decreases the size of the state-space $\mathcal{Y}$, in which case, there will be fewer states for the network to

[1] This means that a node is leak-free when moving from the inactive to the active state but leaky when moving from the active to the inactive state.

visit, improving the likelihood of visiting the zero state and thus reducing the average return time to that state. The confluence of these two effects may contribute to the creation of a noise-induced mode at the state of zero fractional activity.

Equation (5.2.52) shows that, in addition to the net population size N, other system-specific parameters may also influence the mean escape time from the origin. For example, reducing the value of $\lambda_k^+ + \phi_k(\rho_k(\mathbf{0}))$, for every $k \in \mathcal{K}$, will increase the mean time spent at the origin when the LMN reaches complete inactivity.

We should note here that similar arguments can be made to show that, under appropriate conditions, reducing the net population size and tuning system-specific parameters can result in a noise-induced mode at the state $\mathbf{1}$ of maximum fractional activity. Indeed, from Eq. (5.2.51), we have

$$[T_e(\mathbf{1})]^{-1} = N \sum_{k \in \mathcal{K}} \zeta_k \big[\lambda_k^- + \gamma_k(\rho_k(\mathbf{1}))\big]. \tag{5.2.53}$$

This implies that, by reducing N and $\lambda_k^- + \gamma_k(\rho_k(\mathbf{1}))$, for every $k \in \mathcal{K}$, we can increase the mean time spent at $\mathbf{1}$ while improving the likelihood of visiting $\mathbf{1}$ and thus reducing the average return time to this state. As a consequence, the confluence of these two effects may contribute to the creation of a noise-induced mode at the state of maximum fractional activity.

Let us now define a parameter

$$b := \sum_{k \in \mathcal{K}} \zeta_k [\lambda_k^+ + \phi_k(\rho_k(\mathbf{0}))] \geq 0 \tag{5.2.54}$$

and assume that the macroscopic equations (5.2.16) have only one stable fixed point $\boldsymbol{\mu}^*$. Note that, when $b = 0$, the macroscopic equations (5.2.16) imply that the origin of the state-space $\mathcal{Y}$ is also a fixed point (albeit an unstable one) for the macroscopic equations. Thus b is a bifurcation parameter, since the macroscopic equations predict that bifurcation takes place at $b = 0$. Regardless how close the LMN is at the bifurcation point, the macroscopic equations predict that there will be no stable fixed point at the origin, with the dynamics moving away from $\mathbf{0}$ and towards the stable fixed-point $\boldsymbol{\mu}^*$.

On the other hand, our previous discussion implies that, at sufficiently small population sizes, the LMN may behave as if there is a stable fixed point at the origin, in the sense that the network will be operating close to $\mathbf{0}$ with non-negligible probability. This is also true for small nonzero values of the bifurcation parameter b, since Eq. (5.2.52) implies that the mean escape time from the origin is given by $T_e(\mathbf{0}) = (Nb)^{-1}$. This clearly demonstrates that, intrinsic noise present in the network at small population sizes is capable of blurring the bifurcation point from being a single point at $b = 0$ to a small nonnegative neighborhood of 0 while "stabilizing" the unstable fixed point of the macroscopic equations located at the origin of the state-space $\mathcal{Y}$.

Another way to see how intrinsic noise contributes to the creation and stability of a noise-induced mode is by means of the LNA method. For sufficiently large network sizes Ω, the LNA method predicts that the stationary probability distribution $\overline{P}_\Omega(\boldsymbol{y})$ can be sufficiently characterized by a multivariate Gaussian distribution tightly centered around $\boldsymbol{\mu}^*$ with most probability mass being assigned over the state-space $\mathcal{Y}$ which, for all practical purposes, can be thought of as being continuous.[2] As a consequence, the net probability mass assigned by this distribution over the state-space $\mathcal{Y}$ will approximately equal to 1, as expected. However, as the network size Ω decreases, the Gaussian distribution becomes wider around $\boldsymbol{\mu}^*$ and may appreciably extend beyond the state-space $\mathcal{Y}$, which will be discrete for small enough Ω. In this case, the net probability mass assigned at values outside the state-space $\mathcal{Y}$ will not be negligible, and the net probability mass assigned over $\mathcal{Y}$ will be smaller than one. As we explained above, and under appropriate conditions, the Gaussian approximation will begin to break down by placing significant probability mass outside of $\mathcal{Y}$. This may force the stationary distribution $\overline{P}_\Omega(\boldsymbol{y})$ to undergo a qualitative change where the lost probability mass may be thought of as being absorbed at the origin, creating a mode in $\overline{P}_\Omega(\boldsymbol{y})$ at $\mathbf{0}$.

From a potential energy landscape perspective, the width of the potential well associated with the peak of the stationary probability distribution at $\boldsymbol{\mu}^*$ will increase as Ω decreases, whereas

[2] Recall that $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_K$, where $\mathcal{Y}_k := \{0, 1/N_k, \ldots, 1\}$.

its depth will decrease. This behavior is also influenced by other system-specific parameters that control the steady-state solution $\overline{C}$ of the Lyapunov equations (5.2.47). On the other hand, the width and depth of the potential well associated with the peak of the stationary probability distribution at $\mathbf{0}$ will both increase as Ω decreases. If the network parameters are such that the potential well at $\boldsymbol{\mu}^*$ is sufficiently wide and shallow, then a state within this potential well will eventually move towards the potential well at $\mathbf{0}$ with high probability and stay there for an appreciable amount of time before exiting.

The previous discussion provides a clear explanation of the fact that intrinsic noise is an important factor for bursting. In addition to Ω (or N) and $\overline{C}$, this behavior also depends on how far $\boldsymbol{\mu}^*$ is from the origin $\mathbf{0}$, since bursting is clearly better pronounced when $\boldsymbol{\mu}^*$ is further away from the origin. In general however this requires a wider potential well at $\boldsymbol{\mu}^*$. Hence, the network size Ω and any other system-specific parameter that affects the steady-state solution $\boldsymbol{\mu}^*$ of the macroscopic equations

$$\frac{d\mu_k(t)}{dt} = [1-\mu_k(t)]\big[\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{\mu}(t)))\big] - \mu_k(t)\big[\lambda_k^- + \gamma_k(\rho_k(\boldsymbol{\mu}(t)))\big],\ t>0,\ k\in\mathcal{K}, \qquad (5.2.55)$$

initialized by $\boldsymbol{\mu}(0)=\mathbf{0}$ (see Appendix D), and the steady-state solution $\overline{C}$ of the Lyapunov equations (5.2.47) will also affect bursting. This allows the LMN to control bursting by employing alternative strategies.

5.3 Examples

We explored our methods by considering two examples: a stochastic version of a one-dimensional SISa model of Methicillin resistant *staphylococcus aureus* (MRSA) infection [191, 192], with one homogenous population of N individuals, and a two-dimensional stochastic neural network (NN) model with two homogeneous populations of an equal number $N/2$ of *excitatory* and *inhibitory* neurons [19]. We were able with the first example to demonstrate for the first time that avalanching can also occur in epidemiology, even when simple models are used. In the second ex-

ample, we show that the LNA method is not an appropriate tool for explaining the emergence of bursting and avalanching in neural network models. Despite a difference in dimensionality and their functional form, the two examples produce surprisingly similar results.

5.3.1 An epidemiological model

In epidemiology, a common model of disease spreading is the SIS model [7]. According to this model, the n-th individual in a directed weighted network $\mathcal{G}$ of N interacting individuals is assumed to be in one of two states with respect to a disease at time t: *susceptible* (S) and *infected* (I). In this case, the state of the epidemiological system at time t is characterized by a random vector $\boldsymbol{X}(t)$ whose n-th element $X_n(t)$ takes value 1, if the n-th individual is infected, and 0, if she is susceptible. It is assumed that recovery from infection does not confer resistance to the disease with infected individuals becoming susceptible after recovery. For example, bacterial infection is modeled well by the SIS model, since an individual who recovers (e.g., through the use of antibiotics) is susceptible to re-infection.

In the classical SIS model, infection can only be transmitted from an infected to a susceptible individual. Note however that some infections can be acquired from other sources, such as the environment, animals, terror attacks, or self-infection. For example, individuals may be colonized by bacteria and be healthy for long periods of time until the bacteria suddenly seize the opportunity to pathogenically infect the individual. As a consequence, we choose here to discuss a slightly more general version of the SIS model known as the SISa model [193, 194].

Although the SISa model is general enough to describe a variety of infections (e.g., due to illnesses, computer viruses, or social contagions), we focused on a matter of pressing concern to public health: infections due to methicillin resistant *staphylococcus aureus* (MRSA). We use the SISa model as a simple model of MRSA outbreaks in which bacterial infections, due to self-infection, are possible and the number of individuals spreading MRSA infection can be small, in which case stochastic modeling is necessary [195]. MRSA has been studied in confined swine populations, where

more invasive and comprehensive data collection is feasible [192].

We assume that the propensity by which the n-th individual transitions from the susceptible to the infected state depends on a net input $r_n(\boldsymbol{x}) = h_n + \boldsymbol{a}_n^T \boldsymbol{x}$, where $h_n \geq 0$ is the propensity of the individual to become infected regardless of her social contacts and $\boldsymbol{a}_n^T$ is the n-th row of the adjacency matrix $\mathbb{A}$ of the underlying network of infectious social contacts. The element $a_{nn'}$ of the adjacency matrix provides the rate at which the n-th susceptible individual will be infected by the n'-th infected individual and, as such, it is assumed to be nonnegative; i.e., $a_{nn'} \geq 0$, for every $n, n' \in \mathcal{N}$, with $a_{nn} = 0$, for every $n \in \mathcal{N}$. Clearly, $r_n(\boldsymbol{x})$ represents the total infectious influence to the n-th individual. As a consequence, we set $p_n^+(\boldsymbol{x}) = (1 - x_n) r_n(\boldsymbol{x})$. On the other hand, we assume that the propensity of the n-th infected individual to recover is constant, given by ℓ_n^-, which implies that $p_n^-(\boldsymbol{x}) = \ell_n^- x_n$.

To simplify the previous model, we assume one homogeneous population of individuals and study the fraction $Y(t)$ of the population that is infected at time t, given by

$$Y(t) = \frac{1}{N} \sum_{n \in \mathcal{N}} X_n(t). \tag{5.3.56}$$

In this case, one finds that $a_{nn'} = w/N$, for every $n, n' \in \mathcal{N}$ such that $n \neq n'$, whereas $h_n = \eta$, $\ell_n^+ = 0$, and $\ell_n^- = \lambda$, for every $n \in \mathcal{N}$. This implies an all-to-all connectivity, which can be justified by considering the fact that, in small populations, it is always possible for any two individuals to come in contact with each other.

The fractional activity process $\{Y(t), t \geq 0\}$ is Markovian, governed by the following master equation:

$$\frac{\partial P(y;t)}{\partial t} = \pi^+(y - 1/N) P(y - 1/N; t) + \pi^-(y + 1/N) P(y + 1/N; t) - \big[\pi^+(y) + \pi^-(y)\big] P(y;t), \tag{5.3.57}$$

initialized with $P(y;0) = \Delta(y)$, where $P(y;t) := \Pr[Y(t) = y \mid Y(0) = 0]$ and

$$\pi^+(y) = N(1 - y)(\eta + wy) \tag{5.3.58}$$

$$\pi^-(y) = N\lambda y. \tag{5.3.59}$$

Moreover, the macroscopic equation (5.2.55) is given by

$$\frac{d\mu(t)}{dt} = \big[1-\mu(t)\big]\big[\eta + w\mu(t)\big] - \lambda\mu(t), \tag{5.3.60}$$

with initial condition $\mu(0) = 0$, whereas the Lyapunov equation (5.2.47) for the noise variance in the LNA method is given by

$$\frac{dr(t)}{dt} = \big[1-\mu(t)\big]\big[\eta + w\mu(t)\big] + \lambda\mu(t) + 2\big\{w\big[1-\mu(t)\big] - \big[\eta + w\mu(t)\big] - \lambda\big\}r(t), \tag{5.3.61}$$

initialized by $r(0) = 0$. This equation is driven by $\mu(t)$ that solves Eq. (5.3.60). Finally, note from Eq. (5.3.58) that $\pi^+(y) > 0$, when $\eta > 0$, $y < 1$. Therefore, $0 \to 1$ and Proposition 2 in Appendix D implies that $Y(t)$ is irreducible.

Using data from a Danish swine herd, the parameters w and λ have been estimated to take values $w = 0.108$ days^{-1} and $\lambda = 0.0571$ days^{-1} [192]. However, the parameter η could not be reliably estimated from these data. To illustrate the case when infections are rarely contracted (i.e., the case when the environment is relatively clean but not completely free of MRSA), we set $\eta = 10^{-4}$ days^{-1}. From Eq. (5.3.58), note that $\pi^+(y) > 0$, when $\eta > 0$, $y < 1$. Therefore, $0 \to 1$ and Proposition 2 implies that $Y(t)$ is irreducible .

Because the system is one-dimensional, its state space is reasonably sized. It was therefore possible to numerically solve the master equation (5.3.57) for $P_\Omega(y;t)$ using the Krylov subspace approximation (KSA) method implemented by the Expokit software package [32]. To do so, we used a tight tolerance parameter of 10^{-6} and a value $K_0 = 30$ for the dimension of the Krylov subspace. We then employed Eq. (5.2.18) to evaluate the potential energy landscape $V_\Omega(y;t)$ and used the solution to the master equation after 30 years as an approximation to the stationary probability distribution $\overline{P}_\Omega(y)$. From this distribution, we numerically evaluated the internal potential energy $\overline{\mathcal{V}}_\Omega$ and entropy $\overline{\mathcal{S}}_\Omega$ at steady-state. We then calculated the stationary free potential energy according to $\overline{\mathcal{A}}_\Omega = \overline{\mathcal{V}}_\Omega - \overline{\mathcal{S}}_\Omega$. We set $N_0 = 200$, in which case, $\Omega = N/200$. By evaluating $\overline{\mathcal{A}}_\Omega$ for $\Omega = 0.005, 0.01, \ldots, 1$, we computed the stationary pressure $\overline{\mathcal{P}}_\Omega$ using Eq. (5.2.30) and subsequently the bulk modulus $\overline{\mathcal{B}}_\Omega$ using Eq. (5.2.31). We approximated all derivatives with respect to Ω using

backward differences with $\Delta\Omega = 0.005$. When required, we drew sample trajectories from the master equation using the exact Gillespie algorithm [14, 46]. Finally, we numerically solved the macroscopic equation (5.3.60) and the Lyapunov equation (5.3.61) using the stiff 'ode23s' solver in MATLAB® with the default parameters. This resulted in a macroscopic stationary steady state of $\mu^* = 0.4719$.

5.3.2 A neural network model

A model that fits well within our framework has been put forth in the literature to explain biological neural networks [19]. This model is based on an interconnected directed weighted network $\mathcal{G}$ of N neurons in a set $\mathcal{N} = \{1, 2, \ldots, N\}$ that can exist in one of two distinct states: an active state, during which a neuron fires an action potential,[3] and a quiescent state, during which a neuron is at rest. In this case, the state of the neural system at time t is characterized by a random vector $\boldsymbol{X}(t)$ whose n-th element $X_n(t)$ takes value 1, if the n-th neuron is active at time t, and 0 otherwise.

Here, we study the stochastic behavior of a group of interacting neurons embedded within a larger neural network [196]. We assume that, when the n-the neuron is inactive, it is driven to become active by a net input $r_n(\boldsymbol{x}) = h_n + \boldsymbol{a}_n^T\boldsymbol{x}$, where $h_n > 0$ is the external input to the neuron that quantifies the influence of surrounding neurons and external environmental factors, and $\boldsymbol{a}_n^T$ is the n-th row of the adjacency matrix $\mathbb{A}$ of the network. If no external input is present, h_n may be chosen to account for the (small) rate at which a neuron might spontaneously fire independently of the incoming synaptic input. The element $a_{nn'}$ of the adjacency matrix $\mathbb{A}$ provides the synaptic weight from the n'-th to the n-th neuron. If $a_{nn'} > 0$, then the n'-th neuron *excites* the n-th neuron making it more likely to spike, whereas, if $a_{nn'} < 0$, then the n'-th neuron *inhibits* the n-th neuron making it less likely to spike. Finally, $a_{nn'} = 0$ indicates that the n-th neuron has no synaptic input from the n'-th neuron. It is assumed that a neuron does not regulate itself, in which case $a_{nn} = 0$, for $n \in \mathcal{N}$.

The propensity by which the n-th neuron transitions from the quiescent to the active state

[3] The active state includes the accompanying refractory period wherein the neuron is hyperpolarized.

is assumed to monotonically depend on the total synaptic input $r_n(\boldsymbol{x})$ to the neuron by means of a function $f_n(r) = \langle r > 0 \rangle \tanh(r)$, where $\langle r > 0 \rangle$ is the Iverson bracket, taking value 1, if $r > 0$, and 0 otherwise. In this case, $p_n^+(\boldsymbol{x}) = (1 - x_n)\langle r_n(\boldsymbol{x}) > 0 \rangle \tanh(r_n(\boldsymbol{x}))$, where we take $\ell_n^+ = 0$. On the other hand, the propensity of the n-th neuron to transition from the active to the quiescent state is assumed to be a constant ℓ_n^-, regardless of the system state, which implies that $p_n^-(\boldsymbol{x}) = \ell_n^- x_n$.

We simplify the previous model by assuming that the neural network under consideration consists of two homogeneous populations $\mathcal{N}_1$ and $\mathcal{N}_2$ of excitatory and inhibitory neurons, respectively. The fractional activity process $\boldsymbol{Y}(t)$ is now two-dimensional, with elements $Y_1(t)$ and $Y_2(t)$ given by

$$Y_1(t) = \frac{1}{N_1} \sum_{n \in \mathcal{N}_1} X_n(t), \quad Y_2(t) = \frac{1}{N_2} \sum_{n \in \mathcal{N}_2} X_n(t). \tag{5.3.62}$$

Due to homogeneity, we find that $a_{nn'} = w_{11}/N_1$, for every $n, n' \in \mathcal{N}_1$ such that $n \neq n'$, $a_{nn'} = w_{22}/N_2$, for every $n, n' \in \mathcal{N}_2$ such that $n \neq n'$, $a_{nn'} = w_{12}/N_2$, for every $n \in \mathcal{N}_1$, $n' \in \mathcal{N}_2$, and $a_{nn'} = w_{21}/N_1$, for every $n \in \mathcal{N}_2$, $n' \in \mathcal{N}_1$. Moreover, $h_n = \eta_1$, for every $n \in \mathcal{N}_1$, $h_n = \eta_2$, for every $n \in \mathcal{N}_2$, $\ell_n^+ = 0$, $\ell_n^- = \lambda_1$, for every $n \in \mathcal{N}_1$, and $\ell_n^+ = 0$, $\ell_n^- = \lambda_2$, for every $n \in \mathcal{N}_2$. Note that the implied all-to-all connectivity is a common assumption in the neuroscience literature [19, 197]. It is usually justified by noting that some regions of the brain are comprised of neurons that are highly interconnected among themselves.

To simplify matters further, we set $N_1 = N_2 = N/2$, $w_{11} = w_{21} = w_e > 0$, $w_{12} = w_{22} = w_i < 0$, $\lambda_1 = \lambda_2 = \lambda$, $\eta_1 = \eta_2 = \eta$, and $\phi_1(\rho) = \phi_2(\rho) = \phi(\rho) = \langle \rho > 0 \rangle \tanh(\rho)$. This implies that $\rho_1(\boldsymbol{y}) = \rho_2(\boldsymbol{y}) = \rho(\boldsymbol{y}) = w_e y_1 + w_i y_2 + \eta$. In this case, the fractional activity process $\{\boldsymbol{Y}(t), t \geq 0\}$ is

Markovian, governed by the master equation (5.2.13) with propensity functions

$$\pi_1^+(y_1,y_2) = \frac{N}{2}(1-y_1)\langle \eta + w_e y_1 + w_i y_2 > 0\rangle \tanh(\eta + w_e y_1 + w_i y_2) \tag{5.3.63}$$

$$\pi_2^+(y_1,y_2) = \frac{N}{2}(1-y_2)\langle \eta + w_e y_1 + w_i y_2 > 0\rangle \tanh(\eta + w_e y_1 + w_i y_2) \tag{5.3.64}$$

$$\pi_1^-(y_1,y_2) = \frac{N}{2}\lambda y_1 \tag{5.3.65}$$

$$\pi_2^-(y_1,y_2) = \frac{N}{2}\lambda y_2. \tag{5.3.66}$$

Moreover, the macroscopic equations (5.2.16) are found to be

$$\frac{d\mu_1(t)}{dt} = \big[1-\mu_1(t)\big]\langle w_e\mu_1(t) + w_i\mu_2(t) + \eta > 0\rangle \tanh(w_e\mu_1(t) + w_i\mu_2(t) + \eta) - \lambda\mu_1(t) \tag{5.3.67}$$

$$\frac{d\mu_2(t)}{dt} = \big[1-\mu_2(t)\big]\langle w_e\mu_1(t) + w_i\mu_2(t) + \eta > 0\rangle \tanh(w_e\mu_1(t) + w_i\mu_2(t) + \eta) - \lambda\mu_2(t), \tag{5.3.68}$$

initialized by $\mu_1(0) = \mu_2(0) = 0$. Finally, the Lyapunov equations (5.2.47) for the noise correlations in the LNA method can be determined by specifying the diffusion terms as

$$D_1(\mu_1,\mu_2) = (1-\mu_1)\langle w_e\mu_1 + w_i\mu_2 + \eta > 0\rangle \tanh(w_e\mu_1 + w_i\mu_2 + \eta) + \lambda\mu_1 \tag{5.3.69}$$

$$D_2(\mu_1,\mu_2) = (1-\mu_2)\langle w_e\mu_1 + w_i\mu_2 + \eta > 0\rangle \tanh(w_e\mu_1 + w_i\mu_2 + \eta) + \lambda\mu_2, \tag{5.3.70}$$

and the derivatives of the drift terms as

$$A_{11}(\mu_1,\mu_2) = -\lambda + \langle w_e\mu_1 + w_i\mu_2 + \eta > 0\rangle\big\{w_e(1-\mu_1)\big[1-\tanh^2(w_e\mu_1 + w_i\mu_2 + \eta)\big] - \tanh(w_e\mu_1 + w_i\mu_2 + \eta)\big\} \tag{5.3.71}$$

$$A_{12}(\mu_1,\mu_2) = \langle w_e\mu_1 + w_i\mu_2 + \eta > 0\rangle\, w_i(1-\mu_1)\big[1-\tanh^2(w_e\mu_1 + w_i\mu_2 + \eta)\big] \tag{5.3.72}$$

$$A_{21}(\mu_1,\mu_2) = \langle w_e\mu_1 + w_i\mu_2 + \eta > 0\rangle\, w_e(1-\mu_2)\big[1-\tanh^2(w_e\mu_1 + w_i\mu_2 + \eta)\big] \tag{5.3.73}$$

$$A_{22}(\mu_1,\mu_2) = -\lambda + \langle w_e\mu_1 + w_i\mu_2 + \eta > 0\rangle\big\{w_i(1-\mu_2)\big[1-\tanh^2(w_e\mu_1 + w_i\mu_2 + \eta)\big] - \tanh(w_e\mu_1 + w_i\mu_2 + \eta)\big\}. \tag{5.3.74}$$

Note finally that, when $\eta + w_e + w_i > 0$, we have that $\mathbf{0} \to \mathbf{1}$ by following a sequence of state transitions whereby all excitatory neurons become active one-by-one and all inhibitory neurons

become active one-by-one. As a consequence, Proposition 2 in Appendix D implies that $\boldsymbol{Y}(t)$ is irreducible.

By following [19], we set $\lambda = 0.1\,\mathrm{ms}^{-1}$ and $\eta = 0.001$. We also defined two new parameters w_s and w_d, given by

$$w_s := w_e + w_i \quad \text{and} \quad w_d := w_e - w_i. \tag{5.3.75}$$

Note that $w_s < w_d$, since $w_i < 0$. Moreover, for $w_s > -\eta$, $\boldsymbol{Y}(t)$ is irreducible. It turns out that the steady-state solution of the macroscopic equations (5.3.67) and (5.3.68) depends only on w_s, whereas bursting is controlled by the value of w_d. By following [19], we set $w_s = 0.2$ and study the behavior of the neural network for various values of w_d. We will start with $w_d = 0.3$, in which case the network is not balanced (i.e., $w_s \not\ll w_d$), and will proceed to examine the effects of w_d.

To do so, we numerically solved the corresponding master equation for the joint probability distribution $P_\Omega(y_1, y_2; t)$ using the KSA method with tolerance parameter of 10^{-30} and a value $K_0 = 50$ for the dimension of the Krylov subspace. We took the value of the tolerance parameter to be appreciably smaller than in the case of the SISa model in order to effectively deal with the increased dimensionality of the state-space $\mathcal{Y}$. Moreover, we took the value of K_0 to be larger than the one used in the case of the SISa model in order to effectively deal with the increased cardinality of $\mathcal{Y}$. Similarly to the case of the SISa model, we employed Eq. (5.2.18) to evaluate the stationary potential energy landscape $\overline{V}_\Omega(y_1, y_2)$. We set $N_0 = 200$, in which case, $\Omega = N/200$, and used the solution to the master equation at 2,000 ms as an approximation to the stationary probability distribution $\overline{P}_\Omega(y_1, y_2)$, since we noticed that the neural network is approximately at steady-state after that time. By using this distribution, we numerically evaluated the internal potential energy $\overline{\mathcal{V}}_\Omega$ and entropy $\overline{\mathcal{S}}_\Omega$ at steady-state. We then calculated the stationary free potential energy according to $\overline{\mathcal{A}}_\Omega = \overline{\mathcal{V}}_\Omega - \overline{\mathcal{S}}_\Omega$. By evaluating $\overline{\mathcal{A}}_\Omega$ for $\Omega = 0.01, 0.02, \ldots, 1$, we computed the stationary pressure $\overline{\mathcal{P}}_\Omega$ using Eq. (5.2.30) and subsequently the bulk modulus $\overline{\mathcal{B}}_\Omega$ using Eqs. (5.2.31). We approximated all derivatives with respect of Ω using backward differences with $\Delta\Omega = 0.01$.[4] When required, we

[4]Since we are interested in the behavior of a LMN *en route* to the thermodynamic limit, we must take $\Delta\Omega$ to be the

drew sample trajectories from the master equation using the exact Gillespie algorithm. Finally, we numerically solved the macroscopic equations (5.3.67) and (5.3.68) and the corresponding Lyapunov equations using the stiff 'ode23s' solver in MATLAB® with the default parameters, which resulted in a macroscopic stationary steady state of $\boldsymbol{\mu}^* = (\mu_1^*, \mu_2^*)^T = (0.5032, 0.5032)^T$.

5.4 Results

5.4.1 Thermodynamic analysis reveals critical behavior in LMNs

For each of the models discussed in the previous section, we computed the probability distribution $P_\Omega(\boldsymbol{y}; t)$ and the potential energy landscape $V_\Omega(\boldsymbol{y}; t) = -(1/\Omega)\ln[P_\Omega(\boldsymbol{y}; t)/P_\Omega(\boldsymbol{y}_\Omega^*(t); t)]$, parameterized by the network size Ω, where $\boldsymbol{y}_\Omega^*(t)$ is a state at which $P_\Omega(\boldsymbol{y}; t)$ attains its (global) maximum. Figures 5.1–5.4 depict movies of the dynamic evolutions of the stationary potential energy landscapes and probability distributions with respect to decreasing Ω. Moreover, Fig. 5.5 and Fig. 5.6 depict four computed thermodynamic quantities as a function of Ω. The results for the two models are qualitatively identical, despite the fact that the dimensionality of their state spaces are different. Note that the internal and free potential energy plots exhibit a deflection point at network size $\Omega_c = 0.175$ (population size $N_c = 35$), for the SISa model, and at $\Omega_c = 0.49$ ($N_c = 98$), for the NN model, revealing critical behavior. This is also evident from the pressure, which experiences a discontinuity at Ω_c and produces a spike in the bulk modulus. On the other hand, the values of the bulk modulus are very close to zero at all other network sizes. For this reason, we can conclude that the SISa and NN models are *robust* with respect to network size (and hence to variations in the strength of intrinsic noise) away from the critical value Ω_c.

minimum allowable change in network size such that $\Delta N\zeta_k = N_0\Delta\Omega\zeta_k$ is integer valued, for all $k = 1, 2, \ldots, K$, where $\zeta_k := N_k/N$ is fixed. Since $K = 2$ and $N_1 = N_2 = N/2$, we have $\zeta_1 = \zeta_2 = 1/2$ and, therefore, we set $\Delta\Omega = 0.01$.

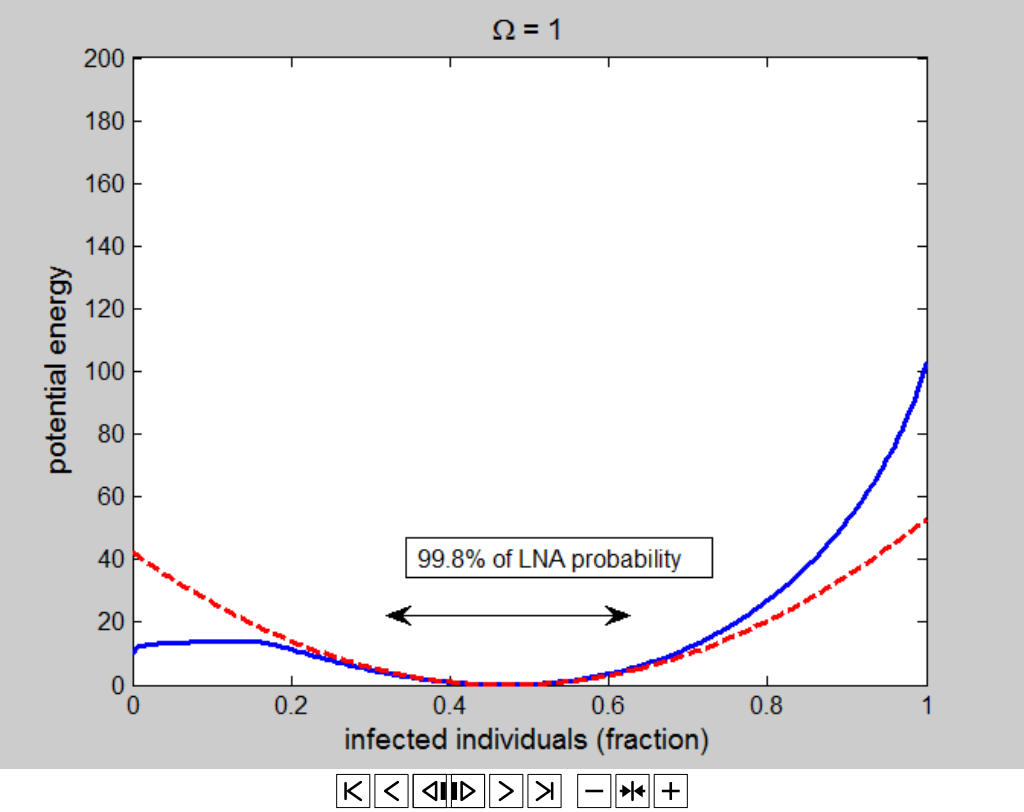

Figure 5.1: Movie of the dynamic evolution, with respect to decreasing network size Ω, of the stationary potential energy landscape of the SISa model (blue solid curve). The red dashed curve represents the potential energy landscape predicted by the LNA method. The double headed arrow indicates the region of 99.8% probability predicted by the LNA method. In addition, the clip indicates when the LNA method produces a probability distribution that extends beyond the state space $\mathcal{Y}$. [May require Adobe Reader (http://get.adobe.com/reader) for proper viewing.]

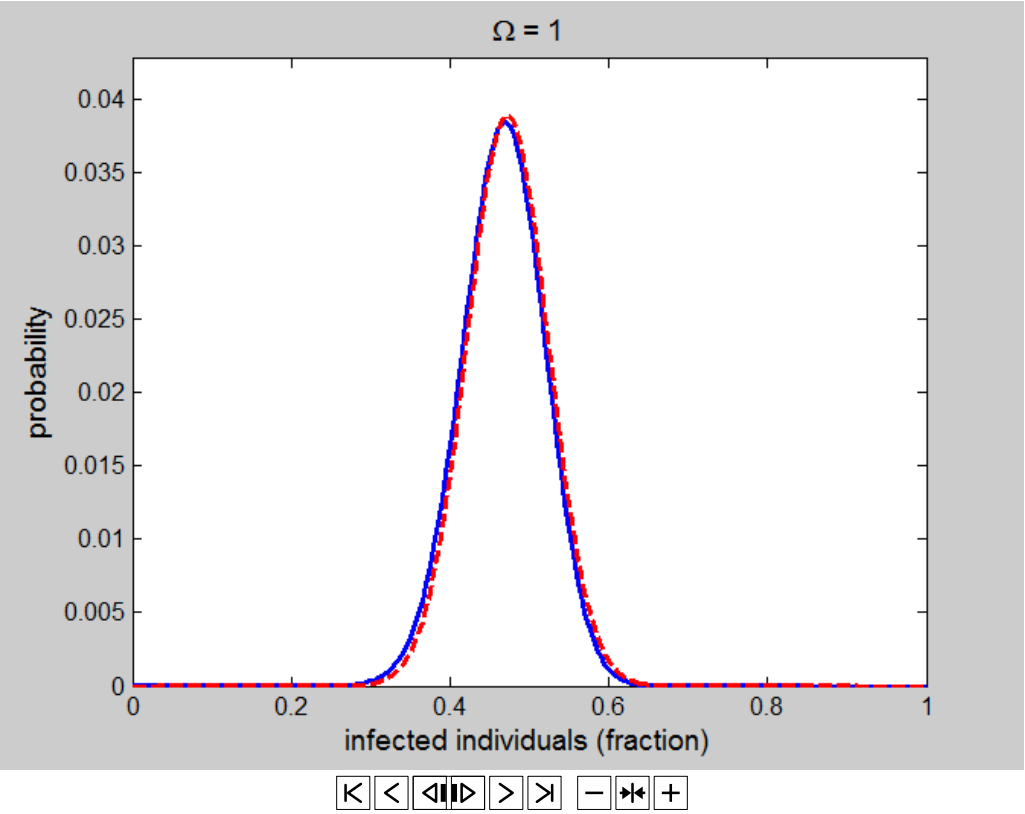

Figure 5.2: Movie of the dynamic evolution, with respect to decreasing network size Ω, of the stationary probability distribution of the SISa model with (blue solid curve). The red dashed curve represents the probability distribution predicted by the LNA method. [May require Adobe Reader (http://get.adobe.com/reader) for proper viewing.]

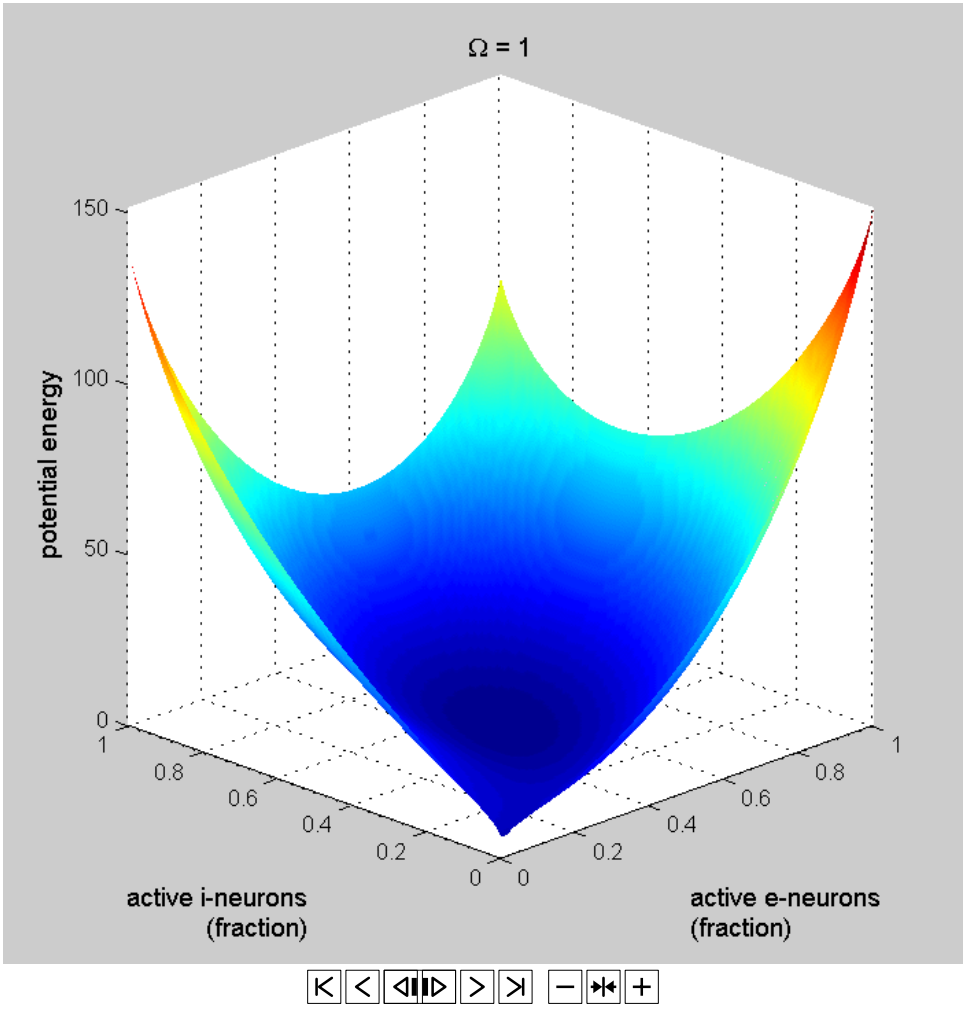

Figure 5.3: Movie of the dynamic evolution, with respect to decreasing network size Ω, of the stationary potential energy landscape of the NN model. [May require Adobe Reader (http://get.adobe.com/reader) for proper viewing.]

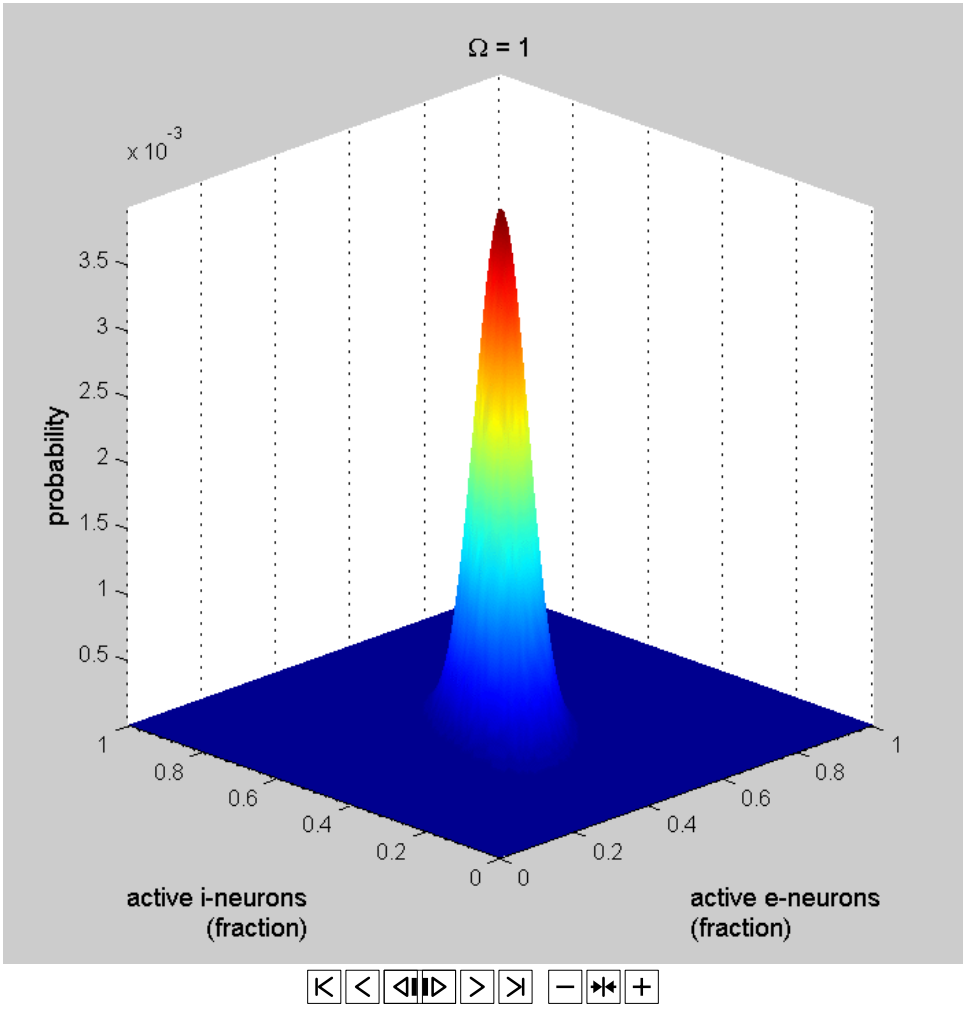

Figure 5.4: Movie of the dynamic evolution, with respect to decreasing network size Ω, of the stationary probability distribution of the NN model. [May require Adobe Reader (http://get.adobe.com/reader) for proper viewing.]

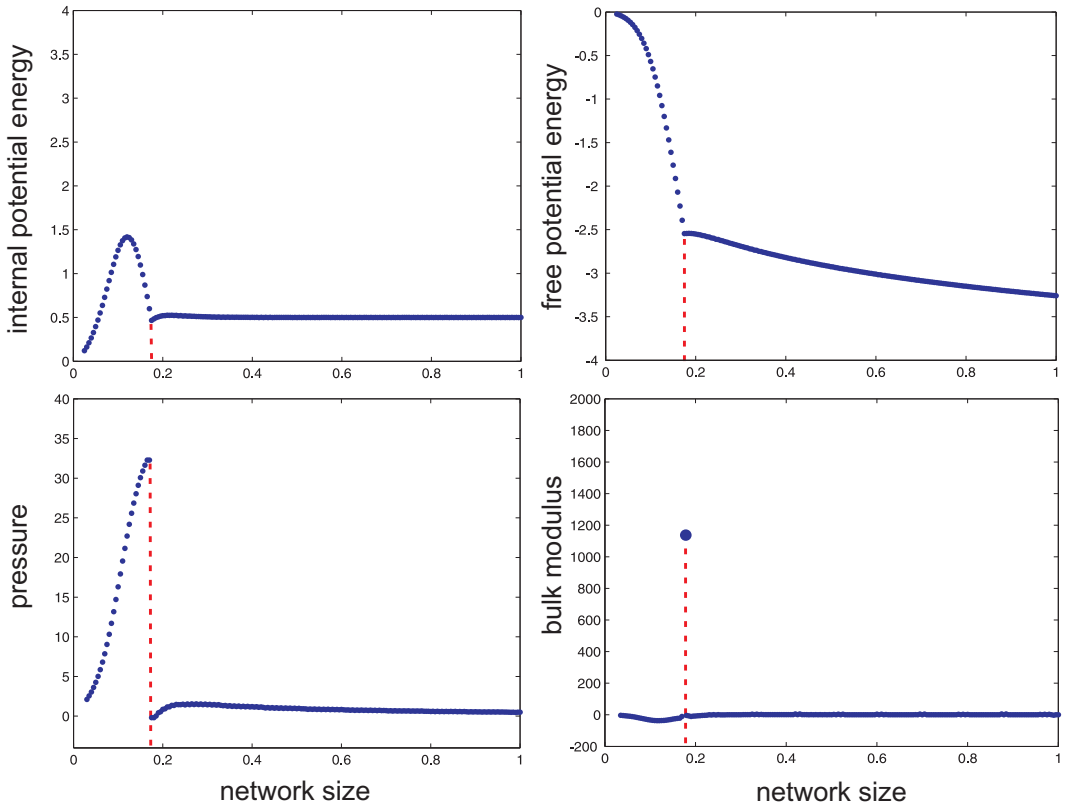

Figure 5.5: Computed thermodynamic quantities for the SISa model as a function of network size Ω. The red dashed lines mark the critical network size $\Omega_c = 0.175$.

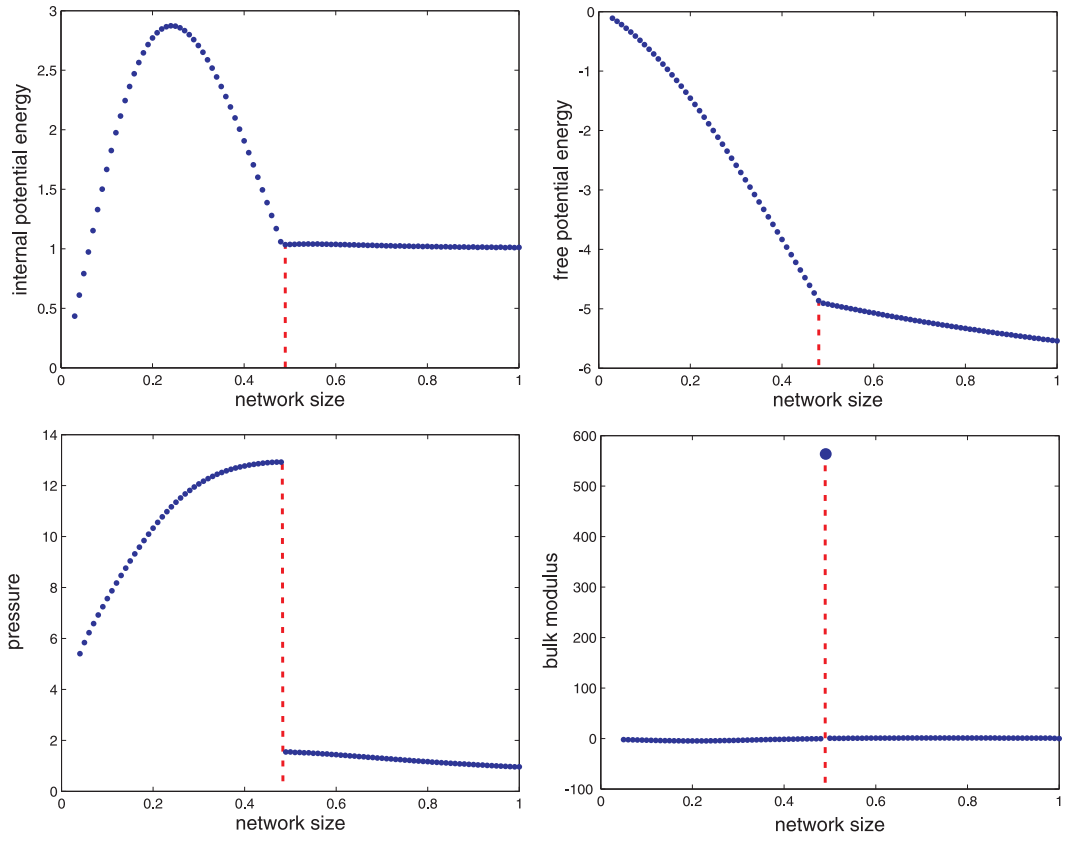

Figure 5.6: Computed thermodynamic quantities for the NN model as a function of network size Ω. The red dashed lines mark the critical network size $\Omega_c = 0.49$.

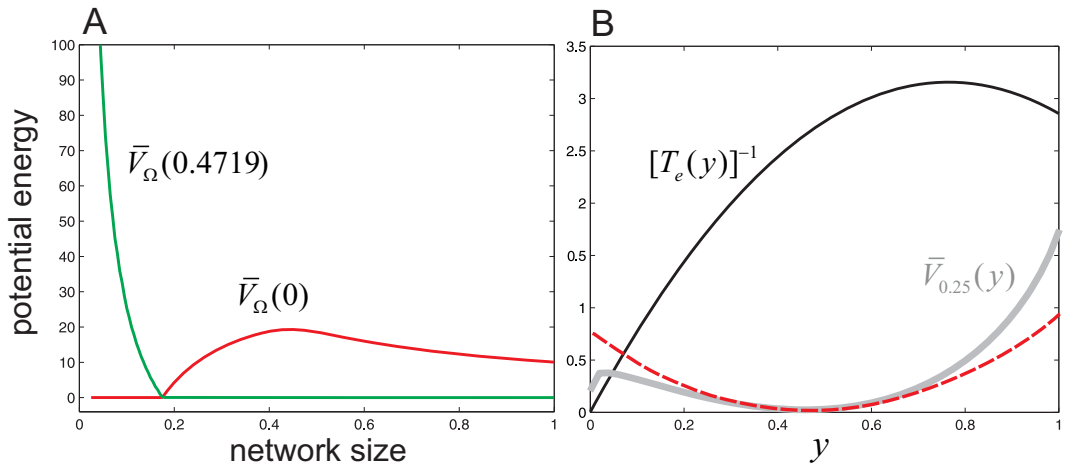

Figure 5.7: (A) Change in the ground state of the potential energy landscape of the SISa model. The red curve depicts $\overline{V}_\Omega(0)$ as a function of Ω, whereas the blue curve depicts $\overline{V}_\Omega(0.4719)$. The two curves intersect at the critical network size $\Omega_c = 0.175$. (B) The inverse mean escape time $[T_e(y)]^{-1}$ from a state y of the SISa model as a function of y, when $\Omega = 0.25$ ($N = 50$), superimposed on the stationary potential energy landscape $\overline{V}_{0.25}(y)$. The red dashed curve depicts the potential energy landscape predicted by the LNA method.

What is the underlying cause of this critical behavior? The previous results suggest that the slope of the self-information support curve $\sigma_*(\Omega) := -\ln \overline{P}_\Omega(\boldsymbol{y}^*_\Omega(\infty); \infty)$ will experience a discontinuity at the critical size Ω_c and a large curvature at that size. This is a consequence of the fact that $\sigma_*(\Omega)$ equals the pressure. Hence, loss of network robustness near Ω_c indicates that there is a change in the ground state (global minimum) of the stationary potential energy landscape at Ω_c. The movies depicted in Fig. 5.1 and Fig. 5.3 corroborate the validity of this point. In particular, Fig. 5.1 confirms that, in the SISa model, critical behavior is caused by the ground state of the potential energy landscape changing from the fixed point 0.4719 of the macroscopic equation to the origin 0 of the state-space as the network size decreases past the critical value $\Omega_c = 0.175$; see also Fig. 5.7A. Likewise, Fig. 5.3 confirms that, in the MM model, critical behavior is caused by the ground state changing from $(0.5032, 0.5032)$ to $(0, 0)$ at $\Omega_c = 0.49$.

5.4.2 LNA fails to accurately predict rare large deviation excursions to the active and inactive states

Figure 5.2 demonstrates that, for large Ω, the LNA method provides a reasonable approximation to the stationary probability distribution of the SISa model. This observation however becomes questionable upon closer examination of the potential energy landscape dynamics depicted in Fig. 5.1. Although the LNA potential energy landscape approximates well the true energy landscape over an appreciable region around the macroscopic ground state $\mu^* = 0.4719$, which accounts for about 99.8% of probability mass, there are substantial differences at the left and right tails of the landscape. These tails characterize *rare* large deviations from the macroscopic ground state and do not conform to the parabolic shape predicted by LNA. As a matter of fact, state values smaller than μ^* reside over a lower and flatter landscape than the one predicted by LNA, whereas state values larger than μ^* reside over a higher and steeper landscape. As a consequence, if the fractional activity process moves to a state at the left end tail of the potential energy landscape (i.e., close to the *inactive* state $y = 0$), it may stay there for an appreciable amount of time before returning back to the macroscopic ground state. On the other hand, if the fractional activity process moves to a state at the right end tail (i.e., close to the *active* state $y = 1$), it may quickly return back to the macroscopic ground state. This behavior, which is not well-predicted by LNA, is corroborated by Fig. 5.7B, which shows the inverse mean escape time $[T_e(y)]^{-1}$ from a state y, as a function of y when $\Omega = 0.25$ ($N = 50$). Similar remarks hold for the NN model.

5.4.3 Stability of the inactive state is directly linked to the strength of intrinsic noise

As the network size Ω decreases towards the critical value Ω_c, the approximation produced by the LNA method begins to break down, due to the emergence of a second well in the potential energy landscape located at the inactive state $\mathbf{0}$; see Fig. 5.1 and Fig. 5.3). This potential well

becomes increasingly dominant, as compared to the well located at $\boldsymbol{\mu}^*$. Since the ground state of the potential energy landscape transitions from $\boldsymbol{\mu}^*$ to $\mathbf{0}$ at the critical size Ω_c and remains at $\mathbf{0}$ for all $\Omega < \Omega_c$, we expect the inactive state to be the most stable state at subcritical network sizes.

The internal potential energy remains fixed at supercritical network sizes; see Fig. 5.5 and Fig. 5.6. This is predicted by Eq. (5.2.49) and the fact that the LNA method provides a good approximation to the solution of the master equation at supercritical sizes (note that $K = 1$ for the SISa model and 2 for the NN model). At subcritical sizes, the internal potential energy monotonically increases initially to a maximum value at some network size Ω_0 ($\Omega_0 = 0.12$ for the SISa model and 0.24 for the NN model) and subsequently monotonically decreases to zero. As a consequence, the internal pressure (which is the derivative of the internal potential energy with respect to Ω) is negative for $\Omega_0 < \Omega < \Omega_c$ and positive for $0 < \Omega < \Omega_0$. Positive internal pressure (decreasing internal potential energy) signifies the fact that removing nodes (individuals or neurons) from the network results in decreasing the distance between the self-information of the most likely state (i.e., the amount of information associated with the occurrence of the inactive state) from the average self-information of all states and thus increasing the stability of this state. As a consequence, and for network sizes below Ω_0, increasing levels of intrinsic noise result in increasing the stability of the inactive state

5.4.4 Emergence of the noise-induced mode leads to bursting

To investigate the emergence of bursting in the SISa model, we depict in the first column of Fig. 5.8 realizations of the fractional activity process (red lines) and the macroscopic dynamics (blue lines), superimposed over the potential energy landscape, for three network sizes, namely $\Omega = 0.25 > \Omega_c$ ($N = 50$), in A, $\Omega = \Omega_c = 0.175$ ($N = 35$), in B, and $\Omega = 0.1 < \Omega_c$ ($N = 20$), in C. Moreover, we depict in the second column of Fig. 5.8 the corresponding stationary potential energy landscapes.

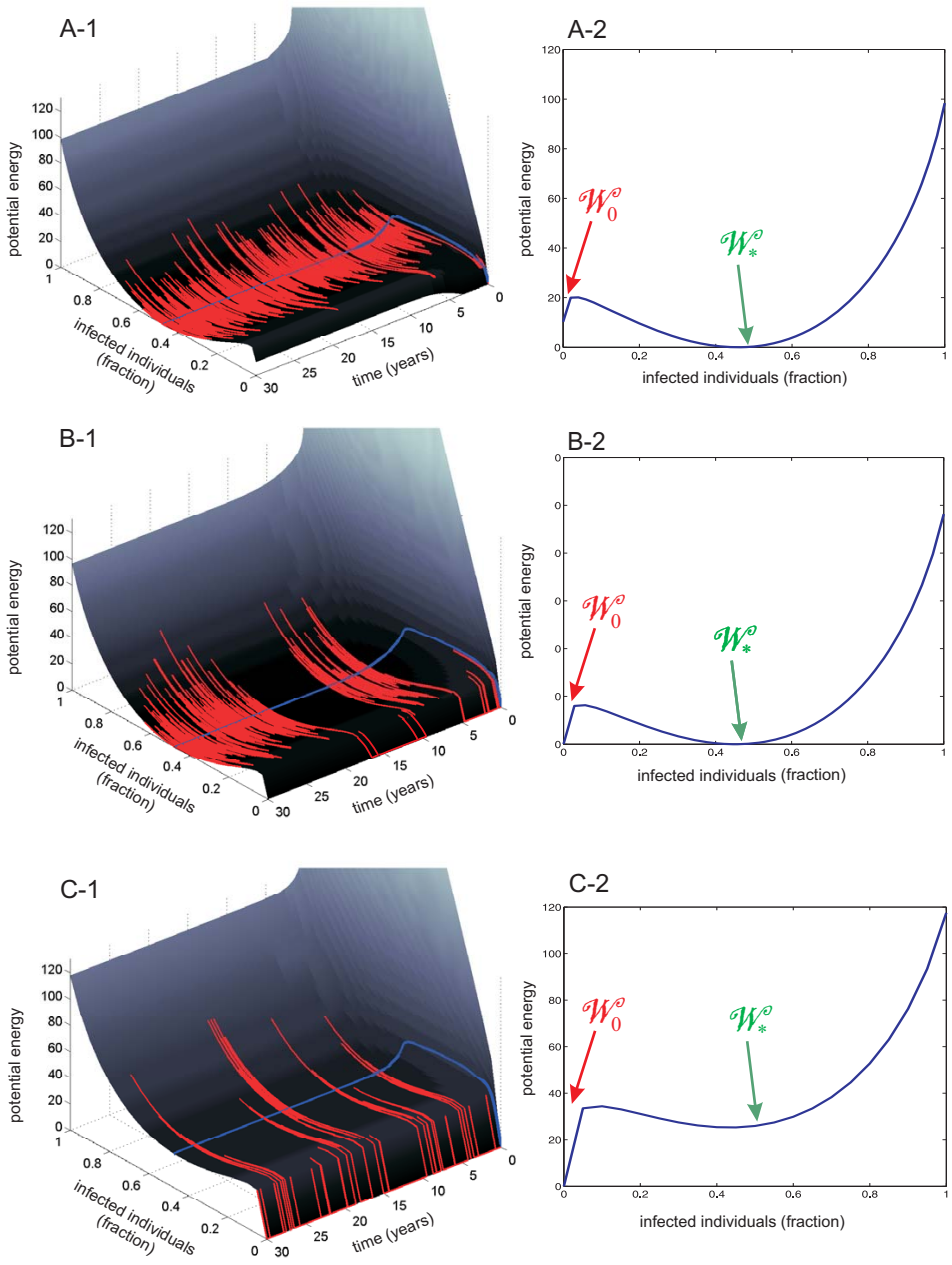

Figure 5.8: Simulations of the SISa model corresponding to: (A) $\Omega = 0.25 > \Omega_c = 0.175$ ($N = 50$), (B) $\Omega = \Omega_c = 0.175$ ($N = 35$), and (C) $\Omega = 0.1 < \Omega_c = 0.175$ ($N = 20$). The left column depicts a single stochastic trajectory of the activity process (in red) along with the corresponding macroscopic solution (in blue), superimposed on the potential energy landscape. The right column depicts the corresponding stationary potential energy landscapes.

The stationary energy landscape depicted in Fig. 5.8A-2 exhibits two potential wells. A shallow and narrow well $\mathcal{W}_0$ located at 0 and a relatively deep and wide well $\mathcal{W}_*$ located at the stable fixed point $\mu^* = 0.4719$ of the macroscopic equation. Transitions from $\mathcal{W}_*$ into $\mathcal{W}_0$ are dubious, since such transitions require appreciable stochastic deviations, which are not likely. On the other hand, transitions from $\mathcal{W}_0$ to $\mathcal{W}_*$ are easier, requiring smaller stochastic fluctuations (mean escape time from 0 is 200 days). In this case, the fraction of infected individuals will fluctuate in a Gaussian-like manner around μ^*, although it may sometimes become zero for a relatively short period of time; see Fig. 5.8A-1.

Figure 5.8B-2 and Fig. 5.8C-2 indicate that, as the network size decreases, the first potential well $\mathcal{W}_0$ becomes deeper and wider, whereas the second well $\mathcal{W}_*$ becomes shallower and eventually disappears; see also Fig. 5.1. When $\Omega = \Omega_c = 0.175$, the two potential wells achieve the same depth. In this case, the fractional activity processes may remain inside $\mathcal{W}_0$ longer than before, since transitions from $\mathcal{W}_0$ into $\mathcal{W}_*$ become more difficult (mean escape time from 0 is now 286 days). As a consequence, the fraction of infected individuals will fluctuate in a Gaussian-like manner around μ^* as before, although it may now become zero for a longer period of time; see Fig. 5.8B-1.

On the other hand, Fig. 5.8C-2 indicates that, when $\Omega = 0.1$, the potential well $\mathcal{W}_*$ becomes extremely shallow. In this case, the fractional activity process will spend most time within $\mathcal{W}_0$ with infrequent and very short excursions outside this well (mean escape time from 0 is 500 days). As a consequence, the fraction of infected individuals will mostly be zero with occasional and brief switching to nonzero values. This bursting behavior is clear from Fig. 5.8C-1 and is expected in the SISa model since, in a hospital setting or in a swine herd, one often speaks of unpredictable "outbreaks" of an infection, such as MRSA. The deterministic SISa model is fundamentally incapable of predicting such complex behavior. Similar remarks apply for the NN model; see Fig. 5.9.

One may be curious about the rather long time scales involved in Fig. 5.8. The infection dynamics occur on the order of months (e.g., an infected individual requires an average time of $1/\lambda = 17.5$ days to recover). However, the long-term infection trends (i.e., outbreaks of infections

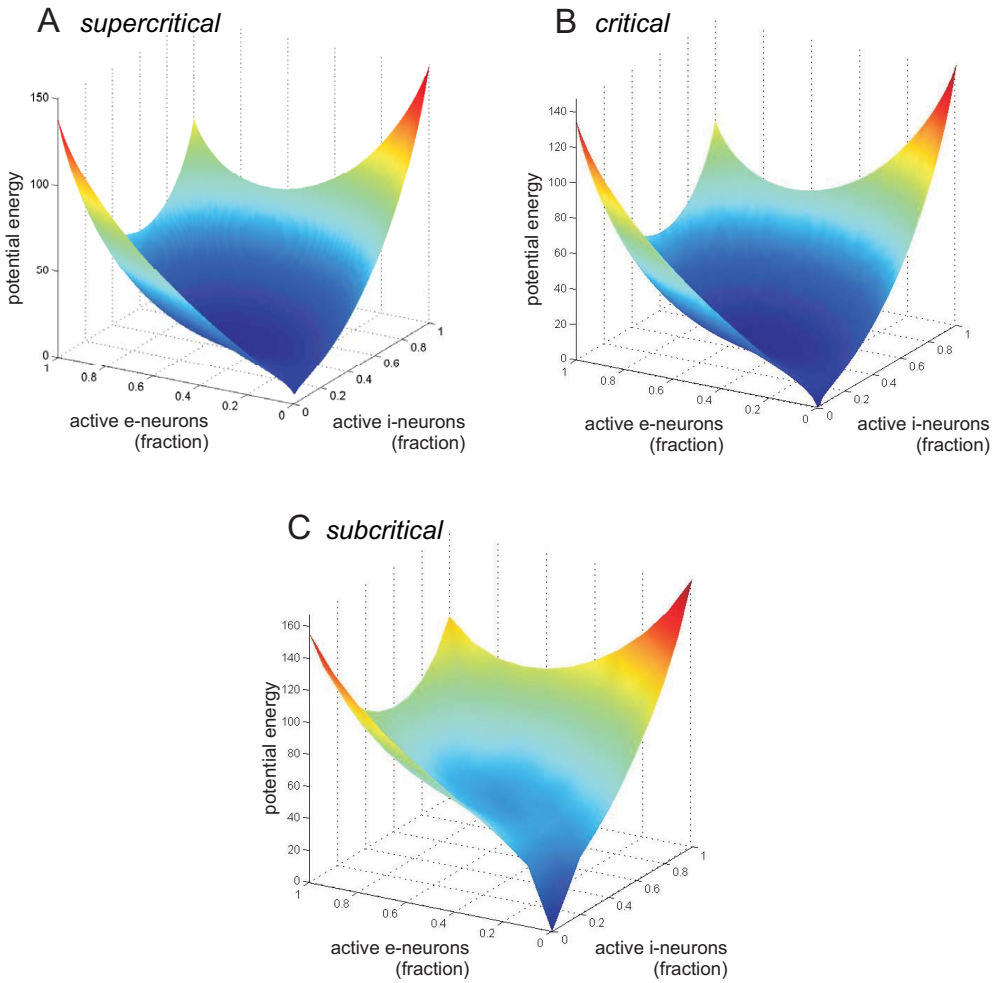

Figure 5.9: Stationary potential energy landscape of the NN model with network size: (A) $\Omega = 0.9$ ($N = 180$); (B) $\Omega = \Omega_c = 0.49$ ($N = 98$), and (C) $\Omega = 0.1$ ($N = 20$).

that eventually die out, only for another outbreak to occur) take place on the time span of multiple years. Recall that the system is initialized with all individuals susceptible and none infected, and therefore nothing happens in the model until an infection is acquired from the environment. This process is captured by the parameter η which is very small, indicating that we are modeling a system where the environment (e.g., the pig pen) is kept relatively clean. Quantitatively, one sees from Eq. (5.3.58) that $\pi^+(0) = N\eta$, and thus the start of a new outbreak is a rare-event for small N and η. Therefore, over a time frame of many years, one may observe multiple outbreaks of MRSA infections.

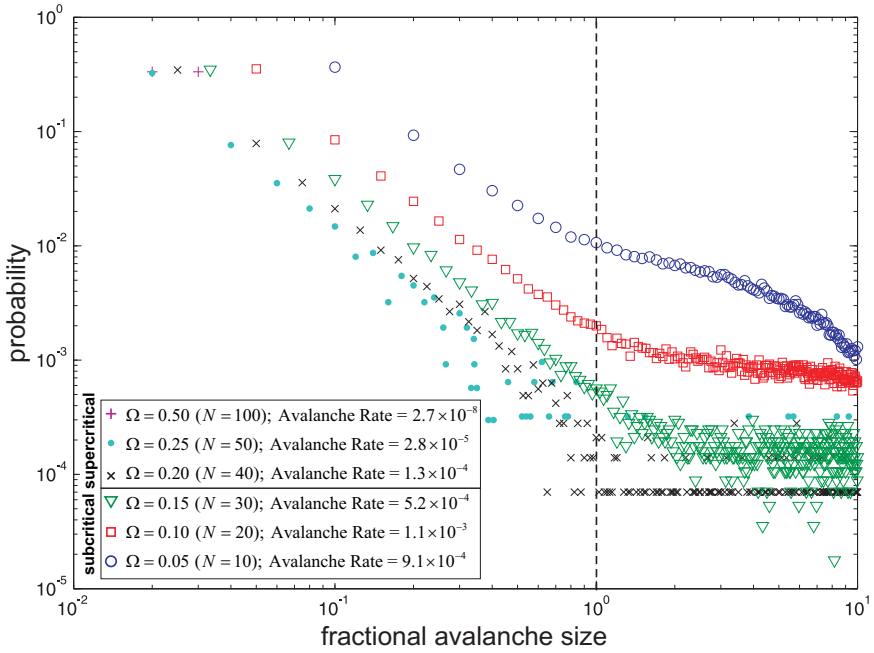

Figure 5.10: Log-log plots of estimated probability distributions of the fractional avalanche size in the SISa model for various network sizes Ω. The cases corresponding to subcritical network sizes below 0.175 exhibit high rates of avalanching with fractional avalanche size distributions characterized by scale-free behavior for sizes smaller than 1. The cases corresponding to supercritical network sizes exhibit increasingly lower rates of avalanching and gradual break-down of scale-free behavior.

5.4.5 Avalanche formation becomes a rare event at supercritical network sizes

Because bursting occurs primarily at steady-state (see Fig. 5.8), we computed avalanche statistics from a single trajectory of the fractional activity process obtained from a long sample of this process (refer to Appendix D to find how we define avalanches). This helped us reduce the computational effort required when calculating avalanche statistics from multiple runs. We simulated the SISa model using the Gillespie algorithm for a period of 300,000 years and used an avalanching threshold $\epsilon = 0.01$ to compute the presence of an avalanche. This allowed us to characterize the SISa model as being active if at least 1 out of 100 individuals was infected. In Fig. 5.10, we depict log-log plots of the estimated probability distributions of the fractional avalanche size, for sizes between 0.01 and 10 and for various choices of Ω. We also depict the rate of avalanche formation for each case, calculated as the number of avalanches that occurred per day. For the three subcritical network

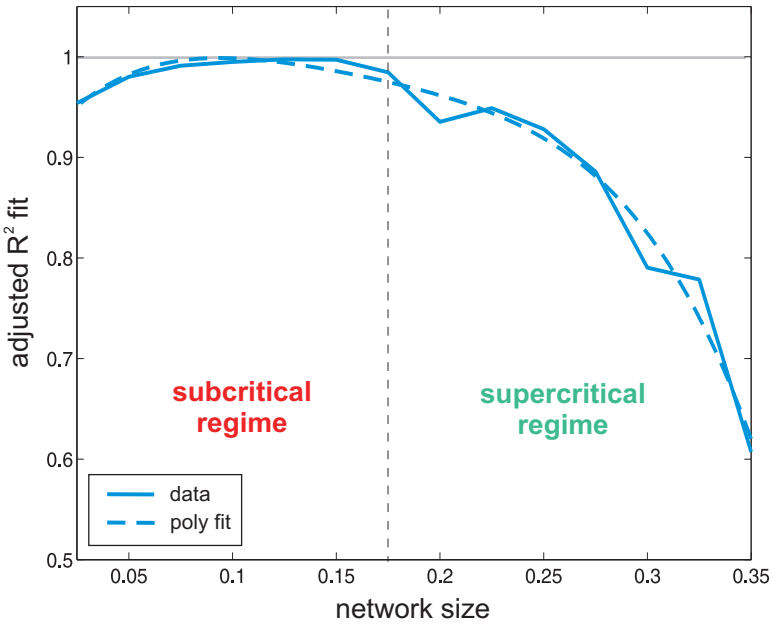

Figure 5.11: Adjusted R^2 values (solid blue curve) of the goodness of fit of a linear regression of a portion (below 1) of the log-log probability distribution of fractional avalanche size for the SISa model computed at discrete network sizes Ω. R^2 values close to one indicate scale-free (linear) behavior. Standard 4-th order polynomial fit of the computed R^2 values produced a smoother curve (dotted blue line). The scale-free property of avalanching is characteristic to network sizes close or below the critical size $\Omega_c = 0.175$ (dotted black curve) and disappears gradually as Ω increases away from the critical size, as indicated by the decreasing R^2 values.

sizes below 0.175, the distributions exhibit scale-free behavior (i.e., the log-log plots are linear) for fractional avalanche sizes below 1 (i.e., when the number of infections that occur during an avalanche is at most N); see also Fig. 5.11. On the other hand, for the three supercritical network sizes above 0.175, we observe increasingly lower rates of avalanching, indicating that avalanche formation becomes eventually a rare event as Ω increases. Moreover, this is accompanied with a loss of the scale-free behavior of the size distribution; see Fig. 5.11. We obtained similar results for the NN model; see Fig. 5.12.

5.4.6 External influences affect bursting

The previous results for the SISa model are based on setting $\eta = 10^{-4}$ days^{-1}. This parameter quantifies the influence of extrinsic factors (other than direct transmission from other

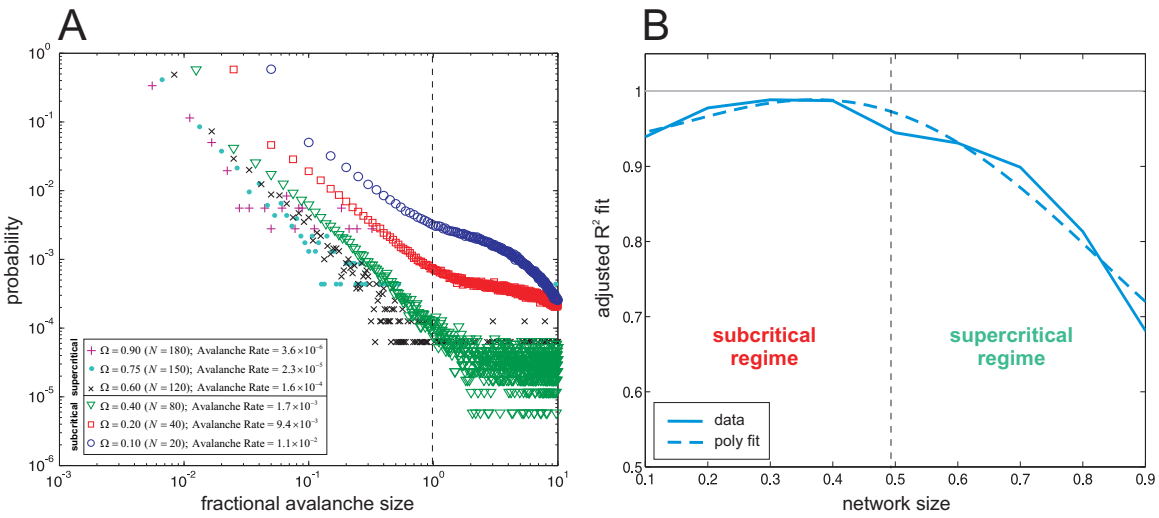

Figure 5.12: (A) Log-log plots of estimated probability distributions of the fractional avalanche size in the NN model for various network sizes Ω. The cases corresponding to subcritical network sizes below $\Omega_c = 0.49$ exhibit high rates of avalanching with fractional avalanche size distributions characterized by scale-free behavior for sizes smaller than 1. The case corresponding to the supercritical network size exhibits a low rate of avalanching and a break-down of the scale-free behavior. (B) Adjusted R^2 values (solid blue curve) of the goodness of fit of a linear regression of a portion (below 1) of the log-log probability distribution of fractional avalanche size for the NN model computed at discrete network sizes Ω. R^2 values close to one indicate scale-free (linear) behavior. Standard 4-th order polynomial fit of the computed R^2 values produced a smoother curve (dotted blue line). The scale-free property of avalanching is characteristic to network sizes close or below the critical size $\Omega_c = 0.49$ (dotted black curve) and disappears gradually as Ω increases away from the critical size, as indicated by the decreasing R^2 values.

infected individuals) on the rate of infection. We therefore investigated the effect of η on bursting.

The mean escape time from the inactive state depends inversely proportional on the network size Ω and parameter η. This implies that, for a fixed value of η, the stability of the inactive state increases for decreasing Ω, in agreement with our previous discussion.

For fixed Ω, $T_e(0) \to \infty$, as $\eta \to 0$, and moving away from the inactive state becomes increasingly difficult. When $\eta = 0$, the SISa model reduces to the standard SIS model of epidemiology, which enjoys far simpler dynamics: infections will always die out and never appear again, since $T_e(0) = \infty$. As a consequence, the stationary probability distribution of the SIS model assigns all probability mass to the inactive state. On the other hand, $T_e(0) \to 0$, as $\eta \to \infty$, which implies that, for sufficiently large η, the SISa model will be moving away from the inactive state almost instantaneously.

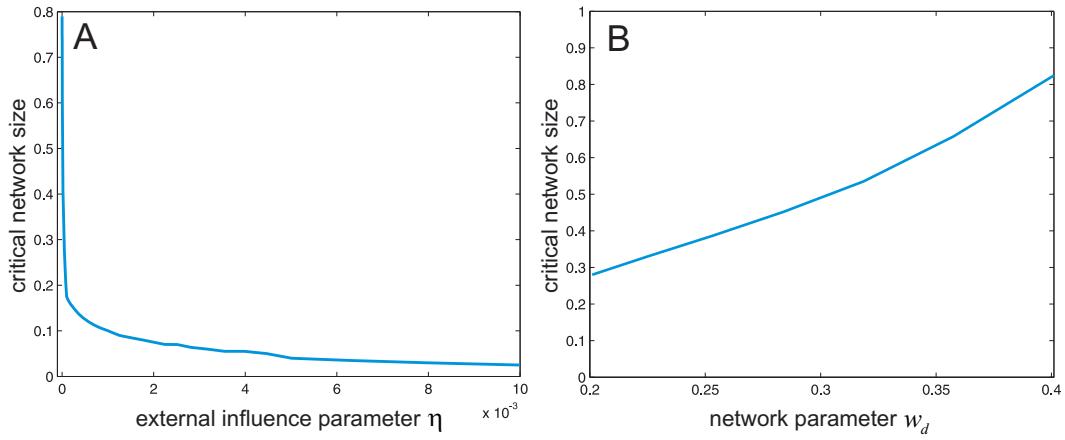

Figure 5.13: (A) Critical network size Ω_c of the SISa model as a function of the external influence parameter η. (B) Critical network size Ω_c of the NN model as a function of parameter w_d.

Figure 5.13A depicts a plot of the critical network size Ω_c as a function of η, for $10^{-5} \leq \eta \leq 10^{-2}$. Clearly, decreasing η increases the value of Ω_c. In particular, $\Omega_c \to \infty$, as $\eta \to 0$. As a consequence, and for sufficiently small values of η, the ground state of the potential energy landscape of the SISa model will be at zero no matter how large Ω is. On the other hand, increasing η decreases Ω_c. In particular, $\Omega_c \to 0$, as $\eta \to \infty$. This implies that, for sufficiently large values of η, the ground state of the potential energy landscape of the SISa model will be at the state μ^* predicted by the macroscopic equation no matter how small Ω is. In this case, bursting will never occur. This is because spontaneous infection from sources other than infected individuals is so prevalent that the state of zero infective individuals has low probability. However, and for the small values of η encountered in practice, the state at 0 will be the ground state for all subcritical network sizes, and bursting behavior will be prevalent. Similar results have been obtained for the NN model.

5.4.7 Balanced feed-forward structure is not necessary for bursting in NNs

In a previous work [19], analysis of the NN model using the LNA method led to the conclusion that for a NN to exhibit bursting it is required that $w_s \ll w_d$, where w_s, w_d are two appropriately defined parameters. When $w_s \ll w_d$, the neural network is *balanced*, in the sense that

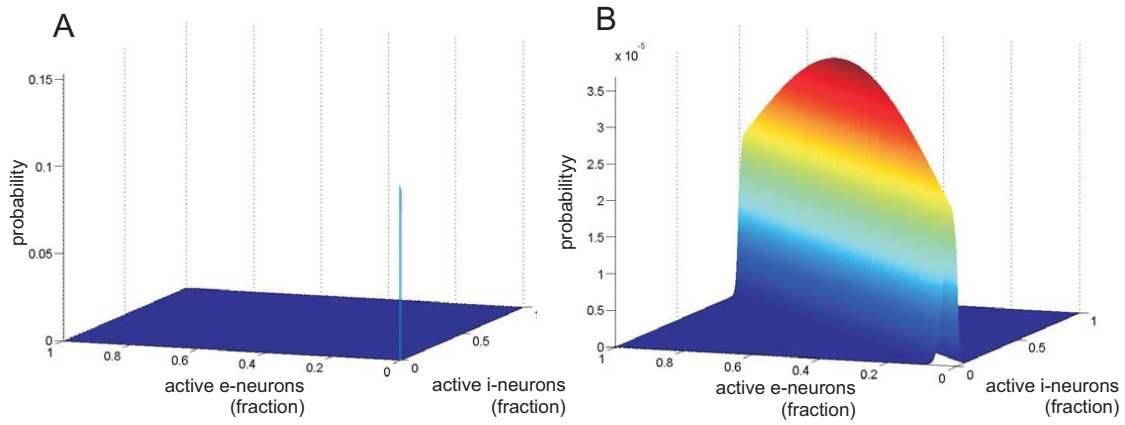

Figure 5.14: (A) The true stationary probability of the fractional activity process in the NN model considered in [Benayoun M, Cowan JD, van Drongelen W, Wallace E (2010) PLoS Comput Biol 6: e1000846], with $N = 1,600$, $\eta = 0.001$, $\lambda = 0.1$, $w_s = 0.2$, and $w_d = 13.8$. (B) The approximating stationary probability distribution obtained by the LNA method. Clearly, the LNA method provides a poor approximation to the actual probability distribution in this case. In particular, the true distribution depicted in A predicts a probability of 0.45 for the network to be at a state close to the inactive state $\mathbf{0}$ and a probability of 10^{-3} for the network to be at a state within a small neighborhood around the macroscopic mode $\boldsymbol{\mu}^*$. On the other hand, the corresponding probabilities predicted by the sampled Gaussian distribution depicted in B are 1.6×10^{-3} and 4×10^{-3}.

excitation is very close to inhibition. Moreover, it has been shown that, when the LNA method is valid, fluctuations in the average difference $[Y_1(t) - Y_2(t)]/2$ of the fractional activity processes of the excitatory and inhibitory neurons feed-forward into the evolution of the average sum $[Y_1(t) + Y_2(t)]/2$. It was then argued that a balanced feed-forward (BFF) structure is necessary for avalanching in relatively large NNs and that this is achieved through amplification of low levels of intrinsic noise.

Our thermodynamic analysis demonstrates that bursting is actually a noise-induced phenomenon that cannot be characterized by the LNA method. This is due to the fact that, at supercritical network sizes, the LNA method may not sufficiently approximate the potential energy landscape in a neighborhood of the inactive state, whereas the method breaks down completely at subcritical network sizes. As a matter of fact, Fig. 5.14 shows that LNA produces a poor approximation to the potential energy landscape close to $\mathbf{0}$ for the model considered in [19]. This is not surprising, since the LNA method always predicts negligible probability for the activity process to reach the inactive state $\mathbf{0}$ [89]. It turns out that the BFF condition is not necessary for bursting in

NNs. Instead, we have argued that bursting is due to the gradual formation of the noise-induced mode at **0** with decreasing network size.

To further confirm this point, note that BFF behavior is controlled by w_d when w_s is held fixed [19]. With $w_s = 0.2$, the analysis in [19] implies that the NN model will exhibit bursting only when $w_d \gg 0.2$. However, our results show that this is also true when $w_d = 0.3$; see Fig. 5.12. In Fig. 5.13B, we depict the computed critical system size Ω_c for a fixed value $w_s = 0.2$ as a function of $w_d > 0.2$. This result demonstrates that, increasing the value of w_d increases the critical network size. Therefore, for a NN with large Ω to exhibit bursting it is required that the value of w_d be sufficiently larger than the value of w_s. This implies that the NN must be balanced. Although the feed-forward condition is not necessary for bursting, it ensures that, in large NNs, the noise induced mode at **0** remains stable.

5.5 Discussion

Energy landscape theory, combined with thermodynamic analysis, leads to a powerful methodology for the analysis of Markovian networks. By introducing leaky Markovian networks, we developed in this chapter an *in silico* approach for understanding the origins of bursting. We have quantified topographic deformations of the energy landscape as a function of network size and showed that bursting is a complex behavior caused by the emergence of noise-induced modes and reallocation of ground states. This led to a novel view of avalanching as a complex behavior that dominates system dynamics at near-critical or subcritical network sizes caused by appreciable levels of intrinsic noise. Future improvements in computer hardware and software will allow our methods to be used in more complicated problems than the ones considered here in an effort to theoretically understand and experimentally evaluate bursting as well as other complex phenomena.

Chapter 6

Conclusion and Outlook

In this thesis, we presented a coherent view of Markovian processes on networks, which would prove useful to scientists across many disciplines. The master equation framework for characterizing such processes is general and powerful, but this comes at a serious cost: the computational burden of simulating and analyzing Markovian processes on networks is usually enormous. Therefore, new solution techniques and analysis methods must be developed in the future, and those which are less general and more tailored to the structure of a specific problem at hand will likely prove most useful in practice.

In this vein, the numerical solution technique developed in Chapter 3 will prove most useful in certain networked systems, such as those encountered in epidemiology, in which the DA process is bounded and lives in a state space that is not much larger than the state space of the corresponding population process. Accordingly, many approximation techniques available in the literature take advantage of structure that is common to many important problems. The LNA method provides a good example of an approximation technique that exploits the monostability and large size of many systems encountered in practice. Therefore, the statistical testing methodology developed in Chapter 4 (or refinements thereof) will prove useful to practicing scientists well into the future. We lastly believe that the work presented in Chapter 5 has successfully demonstrated the practical value

of developing thermodynamic tools for the analysis of Markov processes on networks. Undoubtedly, there is vast untapped potential for advancement in this fledgling field.

As one might expect, there are numerous problems we did not address in this thesis. One important problem is how to deal with systems whose reactions occur on multiple timescales. Such *stiff* systems pose a major challenge when solving the master equation. For a comprehensive review of methods dealing with stiff systems see [5].

Likewise, we focused only on Markovian dynamics. In these systems, the waiting times between occurrences of successive reactions are exponentially distributed. As a consequence, we did not deal with systems in which the waiting reaction times do not follow an exponential distribution. Stochastic processes on such systems fall under the more general purview of semi-Markov processes. Generalization of the present body of work to deal with these type of processes will serve to bring our current developments to an even wider range of scientific disciplines, such as computer science, where non-Markovian Petri nets are commonly used for these purposes [198].

Another important problem which we did not address in this thesis is dealing with random processes on networks with *evolving* topologies. The present framework may be thought of as describing such processes over a timescale on which the network topology is fixed. Note, however, that the generality of the present framework allows for (perhaps inelegant) accommodation of a changing topology. Suppose that there are K reaction networks with differing topologies. We may introduce an artificial species X_{N+1}, which accounts for the different topologies, that takes K values, $1, 2, \ldots, K$, indicating which network topology is presently in use. This is possible to do mathematically by multiplying all propensities of the k-th reaction network by $[X_{N+1} = k]$, where $[\cdot]$ is the Iverson bracket. The multiplication ensures that the reactions can only take place when the "topology" species indicates that those reactions are currently in use. Of course, it is also necessary to include reactions (and the corresponding propensity functions) that modify X_{N+1}, to indicate how the topology changes with time. Although this approach is general enough to accommodate most cases of interest, it comes at the cost of great computational complexity. The system now is

governed by a DA process whose dimensionality grows *multiplicatively* (i.e., if the k-th network has M_k reactions, then the full network requires $\prod_{k=1}^{K} M_k$ reactions to represent all possible topologies, along with a reaction for each possible transition between the K topologies).

Recently, several articles have appeared in the literature introducing *adaptive* networks that explicitly take into account the interplay between network topology and dynamics [199–206]. These preliminary works clearly demonstrate that a number of intriguing properties emerge, not previously observed in nonadaptive networks: formation of complex topologies, spontaneous emergence of modular organization, more complex dynamics than the ones observed in nonadaptive models, and self-organization towards a highly robust critical behavior characterized by power-law distributions.

Another important problem we did not address is sensitivity analysis. Often, the main focus of analysis of the dynamic behavior of a reaction network is a response function that encapsulates some important system characteristics. In epidemiology, for example, one may not care so much about the specific details of the population dynamics, but would rather focus on the total number of individuals infected by a disease over a given period of time. Another example would be the case of cell signaling, where the detailed interactions of a signaling pathway are not as important as the total amount of a protein produced at the "output" of the pathway. Sensitivity analysis is a quantitative approach designed to investigate how variations in the parameters of a reaction network (e.g., in the specific probability rate constants associated with the propensity functions of a mass action system) affect a response function of interest [207–210].

Many physical and man-made reaction networks are designed to be robust to random fluctuations (or even failures) in system components. Although robustness is a highly desirable property, it results in a small number of parameters having a disproportionately large influence on the system response. As a consequence, a robust reaction network can be quite vulnerable to targeted attacks on influential components, which can be a blessing or a curse, depending on the particular situation at hand. For example, development of new drugs may greatly benefit from this property since, to reduce or even eliminate the effects of a disease caused by misregulation of key

system responses, it may be sufficient to design a drug that only inhibits influential reactions that shape these responses. On the other hand, targeted attacks on national infrastructure by hackers or terrorists may produce large scale disruptions with devastating results.

The objective of sensitivity analysis is to determine those factors in a reaction network that produce no noticeable variations in system response and identify those factors that are most influential in shaping that response. Although this is a powerful analysis technique with important practical consequences, it comes with a large computational cost, even in the case of reaction networks with deterministic dynamics [211–213]. For this reason, the development of practical methods for sensitivity analysis of Markovian reaction networks is still in their infancy [144, 214–223].

In the stochastic context, sensitivity analysis involves computing the solution of the master equation using different parameter values. As a consequence, developing efficient solution methods which can be implemented on parallel computer architectures, paired with novel sensitivity estimators, will ensure the feasibility of this type of analysis. Finally, it has been recently demonstrated in [211, 213] that, at least for the case of physical reaction networks with deterministic dynamics, sensitivity analysis methods must be in agreement with underlying thermodynamic constraints. As a consequence, developing accurate, computationally efficient, and thermodynamically consistent sensitivity analysis methods for Markovian reaction networks is an important research activity with significant benefits.

Another fundamental problem for future investigation is finding an appropriate model for a given system by means of statistical inference. In general, there are two fundamentally different types of parameters associated with a Markovian reaction network model: the stoichiometric coefficients ν_{nm} and ν'_{nm} that determine the structure of the network, and the kinetic parameters that determine the non-structural portion of the propensity functions. Some parameter values can be deduced experimentally or by means of appropriate theoretical and sometimes heuristic arguments. Most parameters however must be estimated from available data using statistical inference techniques. Since the predictive power of a given model is fundamentally constrained by the accuracy of its

parameterization, inferring the unknown parameter values in a Markovian reaction network is a problem of paramount interest and practical importance. Although this problem has been extensively studied for reaction networks with deterministic dynamics [224–226], the statistical inference of Markovian reaction networks is largely an open research problem. This problem has been recently investigated in [227–234], but the resulting algorithms do not adequately address important issues, such as curse of dimensionality, thermodynamic consistency, and computational efficiency. These methods have been primarily designed for biochemical reaction networks, but can be easily adopted in other applications with little or no effort.

In most approaches to statistical inference, it is quite common to assume known structural parameters and proceed with estimating the kinetic parameters using noisy and sparse measurements of system dynamics. This problem, known as model calibration, is much easier than the problem of estimating the structural parameters, which is often referred to as model selection.

The two most difficult issues associated with model calibration is the curse of dimensionality and the use of non-convex cost functions which complicate numerical optimization. The curse of dimensionality refers to the fast (exponential) growth of the volume of the parameter space as the number of unknown parameters to be estimated increases. As a consequence, the problem of finding the "best" parameter values becomes difficult when the number of unknown parameters becomes large. This is further exacerbated by the non-convex optimization problem of finding these values, which is computationally difficult to solve in most cases of interest [235]. Therefore, the development of statistical techniques for accurate and computationally efficient model calibration of Markovian reaction networks is an extremely challenging problem. Possible ways to attack this problem are to effectively reduce the number of parameters that must be estimated by incorporating appropriate constraints (e.g., constraints imposed by the fundamental laws of thermodynamics [236–240]) and to identify a smaller set of "influential" parameters whose values must be estimated with sufficient precision (e.g., by employing a sensitivity analysis approach [239]). This reduction in dimensionality must be combined with fast algorithms for solving the master equation, with efficient

optimization methods, and appropriately designed experimental protocols for collecting data with high information content about the values of the unknown parameters [241].

In general, model selection is a more difficult problem. Solving this problem will require development of novel hypothesis testing approaches for comparing between two competing network models (e.g., an originally proposed signaling network and another network obtained by adding new reactions) in a rigorous statistical fashion. This approach however requires that both models are calibrated before compared to each other (e.g., by a likelihood ratio test), which substantially adds to the difficulty of the problem. Another major issue is that more complex models are expected to be more capable of closely matching experimental data, but these models may result in undesirable overfitting. It is therefore necessary to develop methods that appropriately penalize model complexity so that the chosen "optimal" model is the most parsimonious model capable of adequately explaining available data. Finally, all of this must be done while taking into account possible constraints imposed on the structural and kinetic parameters of the network (e.g., by prior knowledge on feasible structural parameter values and by the fundamental laws of thermodynamics).

We hope it is clear that much rewarding work remains to be done in the field of Markovian dynamics on reaction networks. In particular, we feel that future work on solution techniques, thermodynamic analysis, stiffness, evolving topologies, sensitivity analysis, and model estimation, will be at least as rewarding and exciting as the work that has been completed thus far.

Appendix A

Derivation of the master equation

The transition probabilities of a Markov process $\{\boldsymbol{Z}(t), t \geq 0\}$ are constrained by the well-known Chapman-Kolmogorov equations [1]:

$$\Pr[\boldsymbol{Z}(t_{q+1})=\boldsymbol{z}_{q+1}|\,\boldsymbol{Z}(t_{q-1})=\boldsymbol{z}_{q-1}] = \sum_{\boldsymbol{z}_q} \Pr[\boldsymbol{Z}(t_{q+1})=\boldsymbol{z}_{q+1}|\,\boldsymbol{Z}(t_q)=\boldsymbol{z}_q]\,\Pr[\boldsymbol{Z}(t_q)=\boldsymbol{z}_q|\,\boldsymbol{Z}(t_{q-1})=\boldsymbol{z}_{q-1}], \tag{A.1}$$

for every triplet (t_{q-1}, t_q, t_{q+1}) of distinct time points $t_{q-1} < t_q < t_{q+1}$. Given that $\boldsymbol{Z}(t) = \boldsymbol{z}'$, let $T(\boldsymbol{z} \mid \boldsymbol{z}')dt$ be the probability that the (homogeneous) Markov process $\boldsymbol{Z}(t)$ moves only once during the infinitesimally small time interval $[t, t+dt)$ to a new state $\boldsymbol{z} \neq \boldsymbol{z}'$. By convention, we set $T(\boldsymbol{z}' \mid \boldsymbol{z}') = 0$, for every $\boldsymbol{z}'$. Moreover, let $T_0(\boldsymbol{z}')dt$ be the probability that no change of state takes place during $[t, t+dt)$. Then,

$$\Pr[\boldsymbol{Z}(t+dt) = \boldsymbol{z} \mid \boldsymbol{Z}(t) = \boldsymbol{z}'] = \big[T_0(\boldsymbol{z}')dt\big]\Delta(\boldsymbol{z}-\boldsymbol{z}') + T(\boldsymbol{z} \mid \boldsymbol{z}')dt, \tag{A.2}$$

where $\Delta(\boldsymbol{z})$ is the Kronecker delta function and

$$T_0(\boldsymbol{z}')dt = 1 - \sum_{\boldsymbol{z}} T(\boldsymbol{z} \mid \boldsymbol{z}')dt. \tag{A.3}$$

APPENDIX A

If we set $t_{q-1} = 0$, $t_q = t$, and $t_{q+1} = t + dt$, then from Eqs. (A.1)–(A.3) we obtain (by also setting $\boldsymbol{z}_{q-1} = \boldsymbol{0}$, $\boldsymbol{z}_q = \boldsymbol{z}'$, and $\boldsymbol{z}_{q+1} = \boldsymbol{z}$)

$$\begin{aligned}\Pr[\boldsymbol{Z}(t+dt) = \boldsymbol{z} \mid \boldsymbol{Z}(0) = \boldsymbol{0}] \; &= \sum_{\boldsymbol{z}'} \Big[1 - \sum_{\boldsymbol{z}} T(\boldsymbol{z} \mid \boldsymbol{z}')dt\Big] \Delta(\boldsymbol{z} - \boldsymbol{z}') \Pr[\boldsymbol{Z}(t) = \boldsymbol{z}' \mid \boldsymbol{Z}(0) = \boldsymbol{0}] \\ &\quad + \sum_{\boldsymbol{z}'} T(\boldsymbol{z} \mid \boldsymbol{z}') \Pr[\boldsymbol{Z}(t) = \boldsymbol{z}' \mid \boldsymbol{Z}(0) = \boldsymbol{0}]dt \\ &= \Pr[\boldsymbol{Z}(t) = \boldsymbol{z} \mid \boldsymbol{Z}(0) = \boldsymbol{0}] \; - \sum_{\boldsymbol{z}'} T(\boldsymbol{z}' \mid \boldsymbol{z}) \Pr[\boldsymbol{Z}(t) = \boldsymbol{z} \mid \boldsymbol{Z}(0) = \boldsymbol{0}]dt \\ &\quad + \sum_{\boldsymbol{z}'} T(\boldsymbol{z} \mid \boldsymbol{z}') \Pr[\boldsymbol{Z}(t) = \boldsymbol{z}' \mid \boldsymbol{Z}(0) = \boldsymbol{0}]dt,\end{aligned}$$

or

$$\begin{aligned}&\frac{\Pr[\boldsymbol{Z}(t+dt) = \boldsymbol{z} \mid \boldsymbol{Z}(0) = \boldsymbol{0}] - \Pr[\boldsymbol{Z}(t) = \boldsymbol{z} \mid \boldsymbol{Z}(0) = \boldsymbol{0}]}{dt} \\ &\qquad = \sum_{\boldsymbol{z}'} \Big\{ T(\boldsymbol{z} \mid \boldsymbol{z}') \Pr[\boldsymbol{Z}(t) = \boldsymbol{z}' \mid \boldsymbol{Z}(0) = \boldsymbol{0}] - T(\boldsymbol{z}' \mid \boldsymbol{z}) \Pr[\boldsymbol{Z}(t) = \boldsymbol{z} \mid \boldsymbol{Z}(0) = \boldsymbol{0}] \Big\},\end{aligned}$$

which, in the limit as $dt \to 0^+$, leads to

$$\frac{\partial p_{\boldsymbol{z}}(\boldsymbol{z}; t)}{\partial t} = \sum_{\boldsymbol{z}'} \Big\{ T(\boldsymbol{z} \mid \boldsymbol{z}') p_{\boldsymbol{z}}(\boldsymbol{z}'; t) - T(\boldsymbol{z}' \mid \boldsymbol{z}) p_{\boldsymbol{z}}(\boldsymbol{z}; t) \Big\}, \tag{A.4}$$

where $p_{\boldsymbol{z}}(\boldsymbol{z}; t) := \Pr[\boldsymbol{Z}(t) = \boldsymbol{z} \mid \boldsymbol{Z}(0) = \boldsymbol{0}]$. This is a differential form of the Chapman-Kolmogorov equation that is commonly known as the master equation [1].

In the case of Markovian reaction networks, the DA process $\boldsymbol{Z}(t)$ can only be updated based upon the firing of reactions. When the m-th reaction fires within $[t, t + dt)$, the state updates instantaneously according to

$$\boldsymbol{z}(t + dt) = \boldsymbol{z}(t) + \boldsymbol{e}_m, \tag{A.5}$$

where $\boldsymbol{e}_m$ is the m-th column of the $M \times M$ identity matrix. Therefore,

$$T(\boldsymbol{z} \mid \boldsymbol{z}') = \begin{cases} \alpha_m(\boldsymbol{z}')dt, & \text{if } \; \boldsymbol{z} = \boldsymbol{z}' + \boldsymbol{e}_m \\ \\ 0, & \text{otherwise.} \end{cases} \tag{A.6}$$

The master equation (2.1.6) is now a direct consequence of Eqs. (A.4)–(A.6). Finally, we can derive the master equation (2.1.8) by differentiating Eq. (2.1.9) and by using the master equation (2.1.6).

APPENDIX A

Probability of next reaction

The probability of next reaction equals the probability $p_t^0(\tau)$ that no reaction takes place during the time interval $[t, t+\tau)$ multiplied by the conditional probability that the m-th reaction occurs during $[t+\tau, t+\tau+dt)$ given that no reaction occurs within $[t, t+\tau)$. The latter conditional probability is given by $\alpha_m(\boldsymbol{z}(t))dt$, due to the fact that $\boldsymbol{Z}(t+\tau) = \boldsymbol{Z}(t) = \boldsymbol{z}(t)$, since no reaction takes place within $[t, t+\tau)$. Therefore,

$$p_t(\tau, m) = p_t^0(\tau)\alpha_m(\boldsymbol{z}(t)). \tag{A.7}$$

We can divide the time interval $[t, t+\tau)$ into L subintervals of length τ/L, in which case $p_t^0(\tau) = [p_t^0(\tau/L)]^L$. Moreover, and in the limit of large L, $p_t^0(\tau/L) = 1 - \sum_{m\in\mathcal{M}} \alpha_m(\boldsymbol{z}(t))\tau/L$, since $p_t^0(\tau/L)$ is the probability that no reaction will occur during an infinitesimally small time interval of length τ/L. Therefore,

$$p_t^0(\tau) = \lim_{L\to\infty} \left[p_t^0(\tau/L)\right]^L = \lim_{L\to\infty} \left[1 - \tau \sum_{m\in\mathcal{M}} \alpha_m(\boldsymbol{z}(t))/L\right]^L = \exp\left\{-\tau \sum_{m\in\mathcal{M}} \alpha_m(\boldsymbol{z}(t))\right\}, \tag{A.8}$$

where the last equality comes from the definition of the exponential function. Eq. (2.3.9) is now a direct consequence of Eq. (A.7) and Eq. (A.8).

Ω-expansion of the master equation

Let us take $\widetilde{\boldsymbol{z}} = \boldsymbol{\zeta} + \Omega^{-1/2}\boldsymbol{\xi}$ and assume that, for all practical purposes, Ω is large enough so that $\widetilde{\boldsymbol{z}} := \boldsymbol{z}/\Omega$ is continuous-valued. Then, the probability density function $p_{\widetilde{\boldsymbol{z}}}(\widetilde{\boldsymbol{z}}; t)$ of $\widetilde{\boldsymbol{Z}}(t; \Omega)$ satisfies (this density function depends on Ω; however, we do not show this dependance for notational simplicity)

$$p_{\widetilde{\boldsymbol{z}}}(\widetilde{\boldsymbol{z}}; t) = p_{\widetilde{\boldsymbol{z}}}(\boldsymbol{\zeta} + \Omega^{-1/2}\boldsymbol{\xi}; t) = p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}; t),$$

where $p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)$ is the probability density function of the noise component $\boldsymbol{\Xi}(t)$. Moreover,

$$\begin{aligned} p_{\widetilde{\mathbf{Z}}}(\widetilde{\boldsymbol{z}}-\widetilde{\boldsymbol{e}}_m;t) &= p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi}-\widetilde{\boldsymbol{e}}_m;t) \\ &= p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}+\Omega^{-1/2}(\boldsymbol{\xi}-\Omega^{1/2}\widetilde{\boldsymbol{e}}_m);t) \\ &= p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}-\Omega^{1/2}\widetilde{\boldsymbol{e}}_m;t), \end{aligned}$$

where $\widetilde{\boldsymbol{e}}_m := \boldsymbol{e}_m/\Omega$. As a consequence, we have that

$$\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)}{\partial \xi_m} = \frac{\partial p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi};t)}{\partial \xi_m} = \Omega^{-1/2}\frac{\partial p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi};t)}{\partial \widetilde{z}_m},$$

which implies

$$\begin{aligned} \frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)}{\partial t} &= \frac{\partial p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi};t)}{\partial t} + \sum_{m\in\mathcal{M}} \frac{d\zeta_m(t)}{dt}\frac{\partial p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi};t)}{\partial \widetilde{z}_m} \\ &= \frac{\partial p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi};t)}{\partial t} + \Omega^{1/2}\sum_{m\in\mathcal{M}} \frac{d\zeta_m(t)}{dt}\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)}{\partial \xi_m}. \end{aligned} \tag{A.9}$$

From the master equation (2.1.6), we have that

$$\frac{\partial p_{\widetilde{\mathbf{Z}}}(\widetilde{\boldsymbol{z}};t)}{\partial t} = \sum_{m\in\mathcal{M}} \alpha_m(\Omega(\widetilde{\boldsymbol{z}}-\widetilde{\boldsymbol{e}}_m);\Omega)p_{\widetilde{\mathbf{Z}}}(\widetilde{\boldsymbol{z}}-\widetilde{\boldsymbol{e}}_m;t) - \alpha_m(\Omega\widetilde{\boldsymbol{z}};\Omega)p_{\widetilde{\mathbf{Z}}}(\widetilde{\boldsymbol{z}};t),$$

since $\boldsymbol{z} = \Omega\widetilde{\boldsymbol{z}}$, in which case $p_{\mathbf{Z}}(\boldsymbol{z};t) = \Omega^{-1}p_{\widetilde{\mathbf{Z}}}(\boldsymbol{z}/\Omega;t) = \Omega^{-1}p_{\widetilde{\mathbf{Z}}}(\widetilde{\boldsymbol{z}};t)$. This equation, together with Eq. (2.3.19), results in

$$\begin{aligned} &\frac{\partial p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi};t)}{\partial t} = \\ &\sum_{m\in\mathcal{M}} \alpha_m(\Omega(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi}-\widetilde{\boldsymbol{e}}_m);\Omega)p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi}-\widetilde{\boldsymbol{e}}_m;t) - \alpha_m(\Omega(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi});\Omega)p_{\widetilde{\mathbf{Z}}}(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi};t) \\ &= \sum_{m\in\mathcal{M}} \alpha_m(\Omega(\boldsymbol{\zeta}+\Omega^{-1/2}(\boldsymbol{\xi}-\Omega^{1/2}\widetilde{\boldsymbol{e}}_m));\Omega)p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}-\Omega^{1/2}\widetilde{\boldsymbol{e}}_m;t) - \alpha_m(\Omega(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi});\Omega)p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t) \\ &= f(\Omega)\sum_{m\in\mathcal{M}} \widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}(\boldsymbol{\xi}-\Omega^{1/2}\widetilde{\boldsymbol{e}}_m))p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}-\Omega^{1/2}\widetilde{\boldsymbol{e}}_m;t) - \widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t) \\ &\quad + f(\Omega)\,\Omega^{-1}\sum_{m\in\mathcal{M}} \widetilde{\alpha}'_m(\boldsymbol{\zeta}+\Omega^{-1/2}(\boldsymbol{\xi}-\Omega^{1/2}\widetilde{\boldsymbol{e}}_m))p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}-\Omega^{1/2}\widetilde{\boldsymbol{e}}_m;t) - \widetilde{\alpha}'_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t) \\ &= f(\Omega)\sum_{m\in\mathcal{M}} \widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}(\boldsymbol{\xi}-\Omega^{-1/2}\boldsymbol{e}_m))p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}-\Omega^{-1/2}\boldsymbol{e}_m;t) - \widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t) \\ &\quad + f(\Omega)\,\Omega^{-1}\sum_{m\in\mathcal{M}} \widetilde{\alpha}'_m(\boldsymbol{\zeta}+\Omega^{-1/2}(\boldsymbol{\xi}-\Omega^{-1/2}\boldsymbol{e}_m))p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}-\Omega^{-1/2}\boldsymbol{e}_m;t) - \widetilde{\alpha}'_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t). \end{aligned} \tag{A.10}$$

APPENDIX A

Now, by using the Taylor series expansion of $\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}(\boldsymbol{\xi}-\Omega^{-1/2}\boldsymbol{e}_m))p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}-\Omega^{-1/2}\boldsymbol{e}_m;t)$ around $\boldsymbol{\xi}$, given by

$$\begin{aligned}
&\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}(\boldsymbol{\xi}-\Omega^{-1/2}\boldsymbol{e}_m))p_{\boldsymbol{\Xi}}(\boldsymbol{\xi}-\Omega^{-1/2}\boldsymbol{e}_m;t) = \\
&\quad \widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t) - \ \Omega^{-1/2}\frac{\partial[\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m} \\
&\quad + \Omega^{-1}\frac{1}{2}\frac{\partial^2[\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^2} - \ \Omega^{-3/2}\frac{1}{6}\frac{\partial^3[\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^3} \\
&\quad + \Omega^{-2}\frac{1}{24}\frac{\partial^4[\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^4} + \ \mathcal{O}(\Omega^{-5/2}),
\end{aligned}$$

and likewise for $\widetilde{\alpha}'_m$, we have that

$$\begin{aligned}
\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)}{\partial t} - \Omega^{1/2}\sum_{m\in\mathcal{M}}\frac{d\zeta_m(t)}{dt}\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)}{\partial\xi_m} = & - \ \Omega^{-1/2}f(\Omega)\sum_{m\in\mathcal{M}}\frac{\partial[\widetilde{\alpha}_m(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m} \\
& + \ \frac{1}{2}\,\Omega^{-1}f(\Omega)\sum_{m\in\mathcal{M}}\frac{\partial^2[\widetilde{\alpha}_m(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^2} \\
& - \ \frac{1}{6}\,\Omega^{-3/2}f(\Omega)\sum_{m\in\mathcal{M}}\frac{\partial^3[\widetilde{\alpha}_m(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^3} \\
& + \ \frac{1}{24}\,\Omega^{-2}f(\Omega)\sum_{m\in\mathcal{M}}\frac{\partial^4[\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^4} \\
& - \ \Omega^{-3/2}f(\Omega)\sum_{m\in\mathcal{M}}\frac{\partial[\widetilde{\alpha}'_m(\boldsymbol{\zeta}(t)+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m} \\
& + \ \frac{1}{2}\,\Omega^{-2}f(\Omega)\sum_{m\in\mathcal{M}}\frac{\partial^2[\widetilde{\alpha}'_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^2} \\
& + \mathcal{O}(\Omega^{-5/2}),
\end{aligned} \tag{A.11}$$

by virtue of Eq. (A.9) and Eq. (A.10). Then, by setting $\tau = \Omega^{-1}f(\Omega)t$ in Eq. (A.11), we obtain

$$\begin{aligned}
\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)}{\partial\tau} = \ & \Omega^{\frac{1}{2}}\sum_{m\in\mathcal{M}}\frac{d\zeta_m(\tau)}{d\tau}\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)}{\partial\xi_m} - \Omega^{\frac{1}{2}}\sum_{m\in\mathcal{M}}\frac{\partial[\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau)+\Omega^{-\frac{1}{2}}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m} \\
& + \frac{1}{2}\sum_{m\in\mathcal{M}}\frac{\partial^2[\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau)+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m^2} - \frac{\Omega^{-\frac{1}{2}}}{6}\sum_{m\in\mathcal{M}}\frac{\partial^3[\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau)+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m^3} \\
& + \frac{\Omega^{-1}}{24}\sum_{m\in\mathcal{M}}\frac{\partial^4[\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^4} - \Omega^{-\frac{1}{2}}\sum_{m\in\mathcal{M}}\frac{\partial[\widetilde{\alpha}'_m(\boldsymbol{\zeta}(\tau)+\Omega^{-1/2}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m} \\
& + \frac{1}{2}\,\Omega^{-1}\sum_{m\in\mathcal{M}}\frac{\partial^2[\widetilde{\alpha}'_m(\boldsymbol{\zeta}+\Omega^{-\frac{1}{2}}\boldsymbol{\xi})p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};t)]}{\partial\xi_m^2} + \mathcal{O}(\Omega^{-3/2}).
\end{aligned}$$

Moreover, by using the Taylor series expansion of $\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi})$ around $\boldsymbol{\zeta}$, given by

$$\begin{aligned}\widetilde{\alpha}_m(\boldsymbol{\zeta}+\Omega^{-1/2}\boldsymbol{\xi}) \;&= \widetilde{\alpha}_m(\boldsymbol{\zeta}) + \Omega^{-1/2}\sum_{m'\in\mathcal{M}}\xi_{m'}\frac{\partial\widetilde{\alpha}_m(\boldsymbol{\zeta})}{\partial\zeta_{m'}} + \Omega^{-1}\frac{1}{2}\sum_{m'\in\mathcal{M}}\sum_{m''\in\mathcal{M}}\xi_{m'}\xi_{m''}\frac{\partial^2\widetilde{\alpha}_m(\boldsymbol{\zeta})}{\partial\zeta_{m'}\partial\zeta_{m''}} \\ &\quad + \Omega^{-3/2}\frac{1}{6}\sum_{m'\in\mathcal{M}}\sum_{m''\in\mathcal{M}}\sum_{m'''\in\mathcal{M}}\xi_{m'}\xi_{m''}\xi_{m'''}\frac{\partial^3\widetilde{\alpha}_m(\boldsymbol{\zeta})}{\partial\zeta_{m'}\partial\zeta_{m''}\partial\zeta_{m'''}} + \mathcal{O}(\Omega^{-2}),\end{aligned}$$

and likewise for $\widetilde{\alpha}'_m$, we obtain

$$\begin{aligned}\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)}{\partial\tau} \;&= \frac{1}{2}\sum_{m\in\mathcal{M}}\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))\frac{\partial^2 p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)}{\partial\xi_m^2} - \sum_{m\in\mathcal{M}}\sum_{m'\in\mathcal{M}}\frac{\partial\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))}{\partial\zeta_{m'}}\frac{\partial[\xi_{m'}p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m} \\ &\quad + \Omega^{-1/2}\,\frac{1}{2}\left[\sum_{m\in\mathcal{M}}\sum_{m'\in\mathcal{M}}\frac{\partial\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))}{\partial\zeta_{m'}}\frac{\partial^2[\xi_{m'}p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m^2}\right. \\ &\qquad - \sum_{m\in\mathcal{M}}\sum_{m'\in\mathcal{M}}\sum_{m''\in\mathcal{M}}\frac{\partial^2\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))}{\partial\zeta_{m'}\partial\zeta_{m''}}\frac{\partial[\xi_{m'}\xi_{m''}p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m} \\ &\qquad \left. - \frac{1}{3}\sum_{m\in\mathcal{M}}\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))\frac{\partial^3 p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)}{\partial\xi_m^3} - 2\sum_{m\in\mathcal{M}}\widetilde{\alpha}'_m(\boldsymbol{\zeta}(\tau))\frac{\partial p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)}{\partial\xi_m}\right] \\ &\quad - \Omega^{-1}\left[\frac{1}{6}\sum_{m\in\mathcal{M}}\sum_{m'\in\mathcal{M}}\sum_{m''\in\mathcal{M}}\sum_{m'''\in\mathcal{M}}\frac{\partial^3\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))}{\partial\zeta_{m'}\partial\zeta_{m''}\partial\zeta_{m'''}}\frac{\partial[\xi_{m'}\xi_{m''}\xi_{m'''}p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m}\right. \\ &\qquad - \frac{1}{4}\sum_{m\in\mathcal{M}}\sum_{m'\in\mathcal{M}}\sum_{m''\in\mathcal{M}}\frac{\partial^2\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))}{\partial\zeta_{m'}\partial\zeta_{m''}}\frac{\partial^2[\xi_{m'}\xi_{m''}p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m^2} \\ &\qquad + \frac{1}{6}\sum_{m\in\mathcal{M}}\sum_{m'\in\mathcal{M}}\frac{\partial\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))}{\partial\zeta_{m'}}\frac{\partial^3[\xi_{m'}p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m^3} - \frac{1}{24}\sum_{m\in\mathcal{M}}\widetilde{\alpha}_m(\boldsymbol{\zeta}(\tau))\frac{\partial^4 p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)}{\partial\xi_m^4} \\ &\qquad \left. + \sum_{m\in\mathcal{M}}\sum_{m'\in\mathcal{M}}\frac{\partial\widetilde{\alpha}'_m(\boldsymbol{\zeta}(\tau))}{\partial\zeta_{m'}}\frac{\partial[\xi_{m'}p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)]}{\partial\xi_m} - \frac{1}{2}\sum_{m\in\mathcal{M}}\widetilde{\alpha}'_m(\boldsymbol{\zeta}(\tau))\frac{\partial^2 p_{\boldsymbol{\Xi}}(\boldsymbol{\xi};\tau)}{\partial\xi_m^2}\right] \\ &\quad + \mathcal{O}(\Omega^{-3/2}),\end{aligned} \tag{A.12}$$

by virtue of Eq. (2.3.21). By assuming that all terms on the right-hand-side of Eq. (A.12) are negligible for sufficiently large Ω, we obtain the linear Fokker-Planck equation (2.3.22).

APPENDIX A

Mixing coefficients of the stationary master equation solution

Since μ_{ij} is the probability that a Markovian reaction network initialized by the i-th transient state in T will reach a persistent state in P_j at steady-state, we have that

$$\mu_{ij} = \sum_{j' \in P_j} \overline{p}_{i,j'} = \sum_{j' \in P_j} p_{i,j'}(\infty),$$

where $\overline{p}_{i,j}$ is the j-th element of the steady-state vector $\overline{\boldsymbol{p}}_i$, given by Eq. (2.4.14). In this case, we have that

$$\begin{aligned}
\mu_{ij} &= \sum_{j' \in P_j} p_{i,j'}(\infty) \\
&= \sum_{j' \in P_j} \int_0^\infty \frac{dp_{i,j'}(\tau)}{dt}\, d\tau \\
&= \int_0^\infty \sum_{j' \in P_j} \left(\sum_{j'' \in P_j} [\mathbb{P}^{(j)}]_{j'j''} p_{i,j''}(\tau) + \sum_{i' \in T} [\mathbb{T}^{(j)}]_{j'i'} p_{i,i'}(\tau) \right) d\tau \\
&= \int_0^\infty \left(\sum_{j'' \in P_j} \left(\sum_{j' \in P_j} [\mathbb{P}^{(j)}]_{j'j''} \right) p_{i,j''}(\tau) + \sum_{j' \in P_j} \sum_{i' \in T} [\mathbb{T}^{(j)}]_{j'i'} p_{i,i'}(\tau) \right) d\tau \\
&= \int_0^\infty \left(\sum_{j'' \in P_j} \left(\sum_{j' \in P_j} [\mathbb{P}^{(j)}]_{j'j''} \right) p_{i,j''}(\tau) + \sum_{j' \in P_j} \sum_{i' \in T} [\mathbb{T}^{(j)}]_{j'i'} p_{i,i'}(\tau) \right) d\tau \\
&= \int_0^\infty \sum_{j' \in P_j} \sum_{i' \in T} [\mathbb{T}^{(j)}]_{j'i'} p_{i,i'}(\tau)\, d\tau \\
&= \sum_{j' \in P_j} \sum_{i' \in T} [\mathbb{T}^{(j)}]_{j'i'} \int_0^\infty [\exp(\mathbb{T}\tau)]_{i'i}\, d\tau \\
&= \sum_{j' \in P_j} \sum_{i' \in T} [\mathbb{T}^{(j)}]_{j'i'} [-\mathbb{T}^{-1}]_{i'i} \\
&= -\sum_{i' \in T} \sum_{j' \in P_j} [\mathbb{T}^{(j)}]_{j'i'} [\mathbb{T}^{-1}]_{i'i}\,,
\end{aligned}$$

where in the second equality we have used the fact that $p_{i,j'}(0) = 0$, for every $j' \in P_j$, and in the fifth equality we have used the fact that the columns of matrix $\mathbb{P}^{(j)}$ add to zero. This shows Eq. (2.4.16).

APPENDIX A

Proof of Eq. (2.5.9)

From Eq. (2.1.8) and Eq. (2.3.18), the fact that $p_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}};t) = \Omega p_{\boldsymbol{X}}(\Omega\widetilde{\boldsymbol{x}};t)$, and the Taylor series expansion of $\widetilde{\pi}_m(\widetilde{\boldsymbol{x}} - \Omega^{-1}\mathbf{s}_m)\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}} - \Omega^{-1}\mathbf{s}_m)$ around $\widetilde{\boldsymbol{x}}$, we have that

$$\sum_{\boldsymbol{r}}\left(\prod_{n\in\mathcal{N}}\frac{1}{r_n!}\right)\left(-\frac{1}{\Omega}\right)^{\sum_{n\in\mathcal{N}} r_n}\frac{\partial^{\sum_{n\in\mathcal{N}} r_n}}{\partial\widetilde{x}_1^{\,r_1}\cdots\partial\widetilde{x}_N^{\,r_N}}\left[\sum_{m\in\mathcal{M}}\left(\prod_{n\in\mathcal{N}} s_{nm}^{r_n}\right)\widetilde{\pi}_m(\widetilde{\boldsymbol{x}})\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}})\right] = 0,$$

where $\boldsymbol{r}$ is an $N\times 1$ vector with elements r_n that take non-negative integer values. To obtain this equation, we consider large enough Ω so that $\Omega^{-1}\widetilde{\pi}'_m(\boldsymbol{x}/\Omega)$ in Eq. (2.3.18) is negligible. The previous result, together with Eq. (2.5.2) and Eq. (2.5.4), approximately implies that

$$\sum_{m\in\mathcal{M}}\sum_{\boldsymbol{r}}\left(\prod_{n\in\mathcal{N}} s_{nm}^{r_n}\right)\left(\prod_{n\in\mathcal{N}}\frac{1}{r_n!}\right)\frac{\partial^{\sum_{n\in\mathcal{N}} r_n}}{\partial\widetilde{x}_1^{\,r_1}\cdots\partial\widetilde{x}_N^{\,r_N}}\left[\left(-\frac{1}{\Omega}\right)^{\sum_{n\in\mathcal{N}} r_n}\widetilde{\pi}_m(\widetilde{\boldsymbol{x}})\exp\Big\{-\Omega V_0(\widetilde{\boldsymbol{x}})\Big\}\right] = 0. \tag{A.13}$$

By evaluating the derivatives in Eq. (A.13) and by keeping only the terms of $\mathcal{O}(\Omega^0)$, we find

$$\exp\Big\{-\Omega V_0(\widetilde{\boldsymbol{x}})\Big\}\sum_{m\in\mathcal{M}}\widetilde{\pi}_m(\widetilde{\boldsymbol{x}})\sum_{\boldsymbol{r}}\prod_{n\in\mathcal{N}}\frac{1}{r_n!}\left[s_{nm}\frac{\partial V_0(\widetilde{\boldsymbol{x}})}{\partial\widetilde{x}_n}\right]^{r_n} = 0,$$

which implies

$$\sum_{m\in\mathcal{M}}\widetilde{\pi}_m(\widetilde{\boldsymbol{x}})\left[\exp\left\{\sum_{n\in\mathcal{N}} s_{nm}\frac{\partial V_0(\widetilde{\boldsymbol{x}})}{\partial\widetilde{x}_n}\right\} - 1\right] = 0, \tag{A.14}$$

for every state $\widetilde{\boldsymbol{x}}$ such that $V_0(\widetilde{\boldsymbol{x}}) < \infty$ [i.e., for every state $\widetilde{\boldsymbol{x}}$ of nonzero probability]. Finally, Eq. (A.14) implies that

$$\sum_{n\in\mathcal{N}}\frac{\partial V_0(\widetilde{\boldsymbol{x}})}{\partial\widetilde{x}_n}\sum_{m\in\mathcal{M}} s_{nm}\widetilde{\pi}_m(\widetilde{\boldsymbol{x}}) \le 0, \tag{A.15}$$

for every state $\widetilde{\boldsymbol{x}}$ such that $V_0(\widetilde{\boldsymbol{x}}) < \infty$, by virtue of the fact that $\exp\{\boldsymbol{x}^T\boldsymbol{y}\} - 1 \ge \boldsymbol{x}^T\boldsymbol{y}$, for any vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. Equation (A.15), together with the macroscopic equations (2.3.25), shows Eq. (2.5.9).

Proof of Eq. (2.5.10)

From Eq. (2.5.7) and Eq. (2.5.8), we have that

$$\overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) = \frac{\exp\Big\{-\Omega V_0(\widetilde{\boldsymbol{x}})\Big\}\exp\Big\{-V_1(\widetilde{\boldsymbol{x}})\Big\}}{\sum_{\boldsymbol{u}}\exp\Big\{-\Omega V_0(\widetilde{\boldsymbol{u}})\Big\}\exp\Big\{-V_1(\widetilde{\boldsymbol{u}})\Big\}}\,. \tag{A.16}$$

If $\widetilde{\boldsymbol{x}} \in \mathcal{G}_0$, then $V_0(\widetilde{\boldsymbol{x}}) = 0$ and Eq. (A.16) results in

$$\lim_{\Omega\to\infty} \overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) = \frac{1 \times \exp\Big\{-V_1(\widetilde{\boldsymbol{x}})\Big\}}{\sum_{\boldsymbol{u}\in\mathcal{G}_0} 1 \times \exp\Big\{-V_1(\widetilde{\boldsymbol{u}})\Big\} + \sum_{\boldsymbol{u}\notin\mathcal{G}_0} 0 \times \exp\Big\{-V_1(\widetilde{\boldsymbol{u}})\Big\}} = \frac{\exp\Big\{-V_1(\widetilde{\boldsymbol{x}})\Big\}}{\sum_{\boldsymbol{u}\in\mathcal{G}_0} \exp\Big\{-V_1(\widetilde{\boldsymbol{u}})\Big\}},$$

by virtue of the fact that, when $\widetilde{\boldsymbol{x}} \notin \mathcal{G}_0$, $V_0(\widetilde{\boldsymbol{x}}) > 0$, in which case $\lim_{\Omega\to\infty} \exp\Big\{-\Omega V_0(\widetilde{\boldsymbol{x}})\Big\} = 0$. This shows the first part of Eq. (2.5.10).

On the other hand, if $\widetilde{\boldsymbol{x}} \notin \mathcal{G}_0$, then Eq. (A.16) implies that

$$\lim_{\Omega\to\infty} \overline{p}_{\widetilde{\boldsymbol{X}}}(\widetilde{\boldsymbol{x}}) = \frac{0 \times \exp\Big\{-V_1(\widetilde{\boldsymbol{x}})\Big\}}{\sum_{\boldsymbol{u}\in\mathcal{G}_0} 1 \times \exp\Big\{-V_1(\widetilde{\boldsymbol{u}})\Big\} + \sum_{\boldsymbol{u}\notin\mathcal{G}_0} 0 \times \exp\Big\{-V_1(\widetilde{\boldsymbol{u}})\Big\}} = 0\,,$$

which shows the second part of Eq. (2.5.10).

Derivation of thermodynamic balance equations

To derive the balance equations discussed in Section 2.6.1, recall first that we consider a Markovian reaction network comprised of $M/2$ pairs of reversible reactions $(2m-1, 2m)$, $m = 1, 2, \ldots, M/2$. In this case, the master equation (2.1.8) can be written in the following form:

$$\frac{\partial p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\partial t} = \sum_{m=1}^{M/2} \big[\rho_m^+(\boldsymbol{x};t) + \rho_m^-(\boldsymbol{x};t)\big], \quad t > 0, \tag{A.17}$$

where $\rho_m^+(\boldsymbol{x};t) = \pi_{2m-1}(\boldsymbol{x} - \boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x} - \boldsymbol{s}_{2m-1};t) - \pi_{2m}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)$ is the net flux of the forward reaction $2m-1$, whereas, $\rho_m^-(\boldsymbol{x};t) = \pi_{2m}(\boldsymbol{x} + \boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x} + \boldsymbol{s}_{2m-1};t) - \pi_{2m-1}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)$ is the net flux of the reverse reaction $2m$. Note also that

$$\rho_m^+(\boldsymbol{x};t) = -\rho_m^-(\boldsymbol{x} - \boldsymbol{s}_{2m-1};t), \quad \text{for } \ m = 1, 2, \ldots, M/2, \tag{A.18}$$

whereas,

$$\boldsymbol{s}_{2m} = -\boldsymbol{s}_{2m-1}, \quad \text{for } \ m = 1, 2, \ldots, M/2, \tag{A.19}$$

due to the reversibility of the reactions.

From Eq. (2.6.3), Eq. (A.17), Eq. (A.18), and the fact that $\sum_{\boldsymbol{x}\in\mathcal{X}} p_{\boldsymbol{X}}(\boldsymbol{x};t) = 1$, we now

have that

$$
\begin{aligned}
\frac{dS(t)}{dt} &= -\sum_{\boldsymbol{x}\in\mathcal{X}} \frac{\partial p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\partial t} \ln p_{\boldsymbol{X}}(\boldsymbol{x};t) - \sum_{\boldsymbol{x}\in\mathcal{X}} p_{\boldsymbol{X}}(\boldsymbol{x};t) \frac{\partial}{\partial t} \ln p_{\boldsymbol{X}}(\boldsymbol{x};t) \\
&= -\sum_{\boldsymbol{x}\in\mathcal{X}} \sum_{m=1}^{M/2} \Big[\rho_m^+(\boldsymbol{x};t) + \rho_m^-(\boldsymbol{x};t)\Big] \ln p_{\boldsymbol{X}}(\boldsymbol{x};t) \\
&= -\frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \Big[\rho_m^+(\boldsymbol{x};t) + \rho_m^-(\boldsymbol{x};t)\Big] \ln p_{\boldsymbol{X}}(\boldsymbol{x};t) \\
&\quad -\frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \Big[\rho_m^+(\boldsymbol{x};t) + \rho_m^-(\boldsymbol{x};t)\Big] \ln p_{\boldsymbol{X}}(\boldsymbol{x};t) \\
&= -\frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^+(\boldsymbol{x};t) \ln p_{\boldsymbol{X}}(\boldsymbol{x};t) - \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^-(\boldsymbol{x};t) \ln p_{\boldsymbol{X}}(\boldsymbol{x};t) \\
&\quad +\frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^-(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t) \ln \frac{1}{p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)} \\
&\quad +\frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^+(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t) \ln \frac{1}{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)} \\
&= -\frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^+(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)}{p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)} - \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^-(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)}{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)} \\
&= \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^-(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)}{p_{\boldsymbol{X}}(\boldsymbol{x};t)} + \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^+(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)}{p_{\boldsymbol{X}}(\boldsymbol{x};t)} \\
&= \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^-(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)\pi_{2m-1}(\boldsymbol{x})\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})}{p_{\boldsymbol{X}}(\boldsymbol{x};t)\pi_{2m-1}(\boldsymbol{x})\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})} \\
&\quad + \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^+(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)\pi_{2m}(\boldsymbol{x})\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})}{p_{\boldsymbol{X}}(\boldsymbol{x};t)\pi_{2m}(\boldsymbol{x})\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})} \\
&= \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \Bigg[\rho_m^+(\boldsymbol{x};t) \ln \frac{\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)}{\pi_{2m}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)} \\
&\qquad\qquad +\rho_m^-(\boldsymbol{x};t) \ln \frac{\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)}{\pi_{2m-1}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)}\Bigg] \\
&\quad - \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \left[\rho_m^+(\boldsymbol{x};t) \ln \frac{\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})}{\pi_{2m}(\boldsymbol{x})} + \rho_m^-(\boldsymbol{x};t) \ln \frac{\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})}{\pi_{2m-1}(\boldsymbol{x})}\right]. \qquad \text{(A.20)}
\end{aligned}
$$

The entropy balance equation (2.6.5) is now a direct consequence of Eq. (A.20) and Eqs. (2.6.6)–(2.6.8).

Likewise, from Eq. (2.1.8), Eq. (2.6.4), Eq. (A.19), and the fact that $\sum_{\boldsymbol{x}\in\mathcal{X}} p_{\boldsymbol{X}}(\boldsymbol{x};t) = 1$,

we now have that

$$
\begin{aligned}
\frac{dF(t)}{dt} &= \sum_{\boldsymbol{x}\in\mathcal{X}} \frac{\partial p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\partial t} \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})} + \sum_{\boldsymbol{x}\in\mathcal{X}} p_{\boldsymbol{X}}(\boldsymbol{x};t) \frac{\partial}{\partial t} \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})} \\
&= \sum_{\boldsymbol{x}\in\mathcal{X}} \frac{\partial p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\partial t} \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})} \\
&= \sum_{m\in\mathcal{M}} \sum_{\boldsymbol{x}\in\mathcal{X}} [\pi_m(\boldsymbol{x}-\boldsymbol{s}_m) p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_m;t) - \pi_m(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t)] \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})} \\
&= \sum_{m\in\mathcal{M}} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_m(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_m;t)}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_m)} + \sum_{m\in\mathcal{M}} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_m(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}{p_{\boldsymbol{X}}(\boldsymbol{x};t)} \\
&= \sum_{m\in\mathcal{M}} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_m(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_m;t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_m) p_{\boldsymbol{X}}(\boldsymbol{x};t)} \\
&= \frac{1}{2} \sum_{m\in\mathcal{M}} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_m(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_m;t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_m) p_{\boldsymbol{X}}(\boldsymbol{x};t)} \\
&\quad + \frac{1}{2} \sum_{m\in\mathcal{M}} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_m(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_m;t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_m) p_{\boldsymbol{X}}(\boldsymbol{x};t)} \\
&= \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \left[\pi_{2m-1}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1}) p_{\boldsymbol{X}}(\boldsymbol{x};t)} \right. \\
&\qquad \left. + \pi_{2m}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m}) p_{\boldsymbol{X}}(\boldsymbol{x};t)} \right] \\
&= + \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \left[\pi_{2m-1}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1}) p_{\boldsymbol{X}}(\boldsymbol{x};t)} \right. \\
&\qquad \left. + \pi_{2m}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m}) p_{\boldsymbol{X}}(\boldsymbol{x};t)} \right] \\
&= \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1}) p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)} \\
&\qquad - \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_{2m}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1}) p_{\boldsymbol{X}}(\boldsymbol{x};t)}{p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})} \\
&\qquad + \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1}) p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)} \\
&\qquad - \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \pi_{2m-1}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x};t) \ln \frac{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1}) p_{\boldsymbol{X}}(\boldsymbol{x};t)}{p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})} \\
&= \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^+(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)} \\
&\quad + \frac{1}{2} \sum_{m=1}^{M/2} \sum_{\boldsymbol{x}\in\mathcal{X}} \rho_m^-(\boldsymbol{x};t) \ln \frac{p_{\boldsymbol{X}}(\boldsymbol{x};t)\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{m=1}^{M/2}\sum_{\boldsymbol{x}\in\mathcal{X}}\rho_m^+(\boldsymbol{x};t)\ln\frac{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x};t)\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})\pi_{2m}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})\pi_{2m}(\boldsymbol{x})} \\
&\qquad + \frac{1}{2}\sum_{m=1}^{M/2}\sum_{\boldsymbol{x}\in\mathcal{X}}\rho_m^-(\boldsymbol{x};t)\ln\frac{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x};t)\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})\pi_{2m-1}(\boldsymbol{x})}{\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})\pi_{2m-1}(\boldsymbol{x})} \\
&= \frac{1}{2}\sum_{m=1}^{M/2}\sum_{\boldsymbol{x}\in\mathcal{X}}\left[\rho_m^+(\boldsymbol{x};t)\ln\frac{\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})}{\pi_{2m}(\boldsymbol{x})\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}\right. \\
&\qquad\qquad \left. + \rho_m^-(\boldsymbol{x};t)\ln\frac{\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})\overline{p}_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})}{\pi_{2m-1}(\boldsymbol{x})\overline{p}_{\boldsymbol{X}}(\boldsymbol{x})}\right] \\
&\qquad - \frac{1}{2}\sum_{m=1}^{M/2}\sum_{\boldsymbol{x}\in\mathcal{X}}\left[\rho_m^+(\boldsymbol{x};t)\ln\frac{\pi_{2m-1}(\boldsymbol{x}-\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x}-\boldsymbol{s}_{2m-1};t)}{\pi_{2m}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)}\right. \\
&\qquad\qquad \left. + \rho_m^-(\boldsymbol{x};t)\ln\frac{\pi_{2m}(\boldsymbol{x}+\boldsymbol{s}_{2m-1})p_{\boldsymbol{X}}(\boldsymbol{x}+\boldsymbol{s}_{2m-1};t)}{\pi_{2m-1}(\boldsymbol{x})p_{\boldsymbol{X}}(\boldsymbol{x};t)}\right].
\end{aligned}
\tag{A.21}
$$

The balance equation (2.6.10) for the Helmholtz free energy is now a direct consequence of Eq. (A.21) and Eq. (2.6.6), Eq. (2.6.8) and Eq. (2.6.12).

Finally, the balance equation (2.6.11) for the internal energy can be easily derived from Eq. (2.6.4), Eq. (2.6.5), and Eq. (2.6.10).

Appendix B

An example of lexicographic ordering

In this section, We use a simple example to illustrate why lexicographic ordering of the elements of the sample space $\mathcal{Z}$ leads to a lower triangular generator matrix $\mathbb{Q}$ in Eq. (3.2.3).

Let us consider the SIR model and denote by Z_1, Z_2 the DAs of the two reactions S+I $\rightarrow$ 2I and I $\rightarrow$ R, respectively. We will assume that, initially, there are two susceptible individuals, one infected individual, and no recovered individuals; i.e., we will assume that $x_1(0) = 2$, $x_2(0) = 1$, and $x_3(0) = 0$. This implies that $0 \leq Z_1(t) \leq 2$ and $0 \leq Z_2(t) \leq 3$, at any time $t > 0$, which is due to the fact that the first reaction will occur at most two times, after which all individuals will be infected, whereas, the second reaction can occur at most three times, after which all individuals will recover from the infection. In this case, lexicographic ordering of the elements of the two-dimensional sample space $\mathcal{Z}$ results in the following twelve points:

$$
\begin{aligned}
\boldsymbol{z}_1 &= (0,0) \\
\boldsymbol{z}_2 &= (0,1) \\
\boldsymbol{z}_3 &= (0,2) \\
\boldsymbol{z}_4 &= (0,3) \\
\boldsymbol{z}_5 &= (1,0) \\
\boldsymbol{z}_6 &= (1,1) \\
\boldsymbol{z}_7 &= (1,2)
\end{aligned}
$$

APPENDIX B

$$\begin{aligned}
\boldsymbol{z}_8 &= (1,3)\\
\boldsymbol{z}_9 &= (2,0)\\
\boldsymbol{z}_{10} &= (2,1)\\
\boldsymbol{z}_{11} &= (2,2)\\
\boldsymbol{z}_{12} &= (2,3).
\end{aligned}$$

As a consequence, the probability vector $\boldsymbol{q}(t)$ in Eq. (3.2.3) is given by

$$\boldsymbol{q}(t) = \begin{pmatrix}
\Pr[Z_1(t)=0, Z_2(t)=0]\\
\Pr[Z_1(t)=0, Z_2(t)=1]\\
\Pr[Z_1(t)=0, Z_2(t)=2]\\
\Pr[Z_1(t)=0, Z_2(t)=3]\\
\Pr[Z_1(t)=1, Z_2(t)=0]\\
\Pr[Z_1(t)=1, Z_2(t)=1]\\
\Pr[Z_1(t)=1, Z_2(t)=2]\\
\Pr[Z_1(t)=1, Z_2(t)=3]\\
\Pr[Z_1(t)=2, Z_2(t)=0]\\
\Pr[Z_1(t)=2, Z_2(t)=1]\\
\Pr[Z_1(t)=2, Z_2(t)=2]\\
\Pr[Z_1(t)=2, Z_2(t)=3]
\end{pmatrix}. \tag{B.1}$$

Let us now assume that the propensity functions of the two SIR reactions are given by

$$\begin{aligned}
\pi_1(x_1,x_2,x_3) &= k_1 x_1 x_2\\
\pi_2(x_1,x_2,x_3) &= k_2 x_2,
\end{aligned}$$

where k_1 and k_2 are two rate constants and x_1, x_2, and x_3 denote the number of susceptible, infected, and recovered individuals, respectively. Since $x_1(0) = 2$ and $x_2(0) = 1$, Eq. (2.1.4) implies that

$$\begin{aligned}
X_1(t) &= 2 - Z_1(t)\\
X_2(t) &= 1 + Z_1(t) - Z_2(t),
\end{aligned}$$

which, together with Eq. (2.1.7), results in

$$\begin{aligned} \alpha_1(z_1, z_2) &= k_1(2 - z_1)(1 + z_1 - z_2) \\ \alpha_2(z_1, z_2) &= k_2(1 + z_1 - z_2), \end{aligned} \tag{B.2}$$

for $0 \leq z_1 \leq 2$ and $z_2 \leq 1 + z_1$, whereas, $\alpha_1(z_1, z_2) = \alpha_2(z_1, z_2) = 0$, otherwise. As a consequence of Eq. (3.2.3), Eq. (B.1), and Eq. (B.2), the generator matrix $\mathbb{Q}$ is given by

$$\begin{pmatrix} -(2k_1 + k_2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ k_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2k_1 & 0 & 0 & 0 & -2(k_1 + k_2) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2k_2 & -(k_1 + k_2) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2k_1 & 0 & 0 & 0 & -3k_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k_1 & 0 & 0 & 3k_2 & -2k_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2k_2 & -k_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k_2 & 0 \end{pmatrix},$$

which is indeed sparse and lower triangular. Note that states which cannot occur are assigned zero propensities. These states correspond to the zero rows in $\mathbb{Q}$ [e.g., this is true for state $(0, 2)$, which is associated with the third row of $\mathbb{Q}$ and would result in a negative number of -1 infected individuals]. Note also that the non-zero diagonal elements of this matrix are all negative, with the remaining nonzero elements being positive. Finally, each column of $\mathbb{Q}$ sums to zero.

APPENDIX B

Invertibility of $\mathbb{I} - \tau\mathbb{Q}$

We will show that matrix $\mathbb{B} := \mathbb{I} - \tau\mathbb{Q}$ is *invertible*, for any $\tau > 0$. Indeed, for each column k of $\mathbb{B}$, the element b_{kk} is strictly greater than the sum of the absolute values of the remaining elements $b_{k'k}$, $k' \neq k$, since

$$b_{kk} = 1 - \tau q_{kk} = 1 + \tau \sum_{m \in \mathcal{M}} \alpha_m(\boldsymbol{z}_k) > \tau \sum_{m \in \mathcal{M}} \alpha_m(\boldsymbol{z}_k) = \tau \sum_{k' \neq k} q_{k'k} = \sum_{k' \neq k} |b_{k'k}|,$$

for $\tau \neq 0$, by virtue of the fact that $q_{kk} = -\sum_{m \in \mathcal{M}} \alpha_m(\boldsymbol{z}_k)$ and $q_{k'k} = \alpha_m(\boldsymbol{z}_k)$. Thus, $\mathbb{B}$ is invertible according to Theorem 6.1.10 in [242].

The IE method preserves probability vectors

We will now show that, at each iteration j, the IE method produces a probability vector $\widehat{\boldsymbol{q}}(t_j)$ for any step-size τ [i.e., all elements of $\widehat{\boldsymbol{q}}(t_j)$ are nonnegative and sum to one]. Since the initial vector $\widehat{\boldsymbol{q}}(0)$ is taken to be a probability vector, we must only show that, if $\widehat{\boldsymbol{q}}(t_{j-1})$ is a probability vector, then $\widehat{\boldsymbol{q}}(t_j)$ is a probability vector as well.

We will first show that

$$\widehat{q}_k(t_{j-1}) \geq 0 \implies \widehat{q}_k(t_j) \geq 0, \quad \text{for every } k = 1, 2, \ldots, K, \tag{B.3}$$

where $\widehat{q}_k(t_{j-1})$ and $\widehat{q}_k(t_j)$ are the k-th elements of $\widehat{\boldsymbol{q}}(t_{j-1})$ and $\widehat{\boldsymbol{q}}(t_j)$, respectively. Note that the off-diagonal elements of matrix $\mathbb{B} := \mathbb{I} - \tau\mathbb{Q}$ are nonpositive, since $b_{k'k} = -\tau q_{k'k} = -\tau\alpha_m(\boldsymbol{z}_k) \leq 0$, for $k' \neq k$. Furthermore, using the same argument as before, we can show that $\mathbb{B} + t\mathbb{I}$ is non-singular for every $t \geq 0$. According to Theorem 2.5.3 in [243], all elements of matrix $\mathbb{B}^{-1}$ are nonnegative. Since $\widehat{\boldsymbol{q}}(t_j) = \mathbb{B}^{-1}\widehat{\boldsymbol{q}}(t_{j-1})$, we obtain Eq. (B.3).

We will now show that

$$\mathbf{1}^T\widehat{\boldsymbol{q}}(t_{j-1}) = 1 \implies \mathbf{1}^T\widehat{\boldsymbol{q}}(t_j) = 1,$$

where the elements of vector $\mathbf{1}$ are all equal to one. Indeed, we have that

$$\begin{aligned} 1 &= \mathbf{1}^T \widehat{\boldsymbol{q}}(t_{j-1}) \\ &= \mathbf{1}^T (\mathbb{I} - \tau\mathbb{Q})\, \widehat{\boldsymbol{q}}(t_j) \\ &= \mathbf{1}^T \widehat{\boldsymbol{q}}(t_j) - \tau \mathbf{1}^T \mathbb{Q} \widehat{\boldsymbol{q}}(t_j) \\ &= \mathbf{1}^T \widehat{\boldsymbol{q}}(t_j) - \tau \mathbf{0}^T \widehat{\boldsymbol{q}}(t_j) \\ &= \mathbf{1}^T \widehat{\boldsymbol{q}}(t_j), \end{aligned} \tag{B.4}$$

where the elements of vector $\mathbf{0}$ are all equal to zero. The second equality in Eq. (B.4) comes from Eq. (3.2.4), whereas, the fourth equality comes from the fact that the elements of each column of matrix $\mathbb{Q}$ sum to zero.

Note that the previous arguments do not depend on the particular value of the step-size τ. Hence, $\widehat{\boldsymbol{q}}(t_j)$ is a probability vector for any value of τ.

Global error of the IE method

In this section, we show that the global error $||\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j)||_1$ associated with the IE method, where $t_j := j\tau$, is of $\mathcal{O}(\tau)$.

Note that $\boldsymbol{q}(t_j) = \exp(\tau\mathbb{Q})\boldsymbol{q}(t_{j-1})$ and $(\mathbb{I} - \tau\mathbb{Q})\widehat{\boldsymbol{q}}(t_j) = \widehat{\boldsymbol{q}}(t_{j-1})$. Thus,

$$\boldsymbol{q}(t_j) = \exp(\tau\mathbb{Q})\boldsymbol{q}(t_{j-1}) = \cdots = \exp(j\tau\mathbb{Q})\boldsymbol{q}(0)$$

$$\widehat{\boldsymbol{q}}(t_j) = (\mathbb{I} - \tau\mathbb{Q})^{-1}\widehat{\boldsymbol{q}}(t_{j-1}) = \cdots = (\mathbb{I} - \tau\mathbb{Q})^{-j}\widehat{\boldsymbol{q}}(0).$$

As a result,

$$\exp(-j\tau\mathbb{Q})\boldsymbol{q}(t_j) = (\mathbb{I} - \tau\mathbb{Q})^j \widehat{\boldsymbol{q}}(t_j), \tag{B.5}$$

since $\widehat{\boldsymbol{q}}(0) = \boldsymbol{q}(0)$. However,

$$\exp(-\tau\mathbb{Q}) = \mathbb{I} - \tau\mathbb{Q} + \mathcal{O}(\tau^2). \tag{B.6}$$

APPENDIX B

From Eq. (B.5) and Eq. (B.6), we have that

$$\exp(-j\tau\mathbb{Q})\boldsymbol{q}(t_j) = \left[\exp(-\tau\mathbb{Q}) - \mathcal{O}(\tau^2)\right]^j \widehat{\boldsymbol{q}}(t_j) = \left[\exp(-j\tau\mathbb{Q}) - j\mathcal{O}(\tau^2)\right] \widehat{\boldsymbol{q}}(t_j).$$

Consequently,

$$\exp(-j\tau\mathbb{Q})[\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j)] = -j\mathcal{O}(\tau^2)\widehat{\boldsymbol{q}}(t_j) \quad \Longleftrightarrow \quad \boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j) = -j\mathcal{O}(\tau^2)\exp(j\tau\mathbb{Q})\widehat{\boldsymbol{q}}(t_j),$$

which implies that

$$\begin{aligned} ||\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j)||_1 &= ||j\mathcal{O}(\tau^2)\exp(j\tau\mathbb{Q})\widehat{\boldsymbol{q}}(t_j)||_1 \\ &\leq j\mathcal{O}(\tau^2)\,||\exp(j\tau\mathbb{Q})||_1||\widehat{\boldsymbol{q}}(t_j)||_1 \\ &= j\mathcal{O}(\tau^2) \\ &= \frac{t_j}{\tau}\,\mathcal{O}(\tau^2) \\ &\leq \frac{t_{\max}}{\tau}\,\mathcal{O}(\tau^2), \end{aligned} \tag{B.7}$$

where $t_{\max}$ is the maximum simulation time. To obtain Eq. (B.7), we have used the fact that $||\exp(j\tau\mathbb{Q})||_1 = 1$, since $\mathbb{Q}$ is the generator matrix of a Markov process, and $||\widehat{\boldsymbol{q}}(t_j)||_1 = 1$. As a result, we finally obtain $||\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j)||_1 \leq t_{\max}\mathcal{O}(\tau)$, which implies that $||\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j)||_1 = \mathcal{O}(\tau)$.

Computational cost of the KSA method

The Arnoldi procedure performed at each step of the KSA method requires L_0 matrix-vector multiplications between matrix $\mathbb{P}$ and the probability distribution $\boldsymbol{p}$, resulting in a cost of $\mathcal{O}(L_0L^2)$ computations in general. However, the sparsity of $\mathbb{P}$ [matrix $\mathbb{P}$ has $(M+1)L$ non-zero elements instead of L^2] reduces this cost to $\mathcal{O}(L_0(M+1)L)$. Additionally, the orthonormalization step in the Arnoldi procedure requires $\mathcal{O}(L_0^2L)$ operations due to inner product computations. Finally, the Krylov subspace approximation step requires that matrix $\mathbb{V}$ is multiplied with the first column of the matrix exponential $\exp(\tau\mathbb{H})$, at a cost of $\mathcal{O}(L_0L)$. By summing these costs, we can see that the total computational cost of the KSA method is of $\mathcal{O}(L_0(M+L_0)L)$. On the other hand, the storage

requirements are of $\mathcal{O}((M+L_0)L)$, where $\mathcal{O}(ML)$ memory locations are required for storing $\mathbb{P}$ and $\mathcal{O}(L_0L)$ locations are required for storing matrix $\mathbb{V}$, which is multiplied with the first column of the matrix exponential $\exp(\tau\mathbb{H})$.

Error of the Richardson extrapolation procedure

To justify the Richardson extrapolation procedure used to improve the accuracy of the IE method, let us assume that the solution $\boldsymbol{q}(t_{j-1})$ of Eq. (3.2.3) is known at time t_{j-1}. Then, the approximate solution $\widehat{\boldsymbol{q}}(t_j \mid t_{j-1})$ obtained by the IE method at time t_j satisfies

$$\widehat{\boldsymbol{q}}(t_j \mid t_{j-1}) = \boldsymbol{q}(t_{j-1}) + \tau\mathbb{Q}\widehat{\boldsymbol{q}}(t_j \mid t_{j-1}), \tag{B.8}$$

by virtue of Eq. (3.2.4). We now have that

$$\begin{aligned}
\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}(t_j \mid t_{j-1}) &= \boldsymbol{q}(t_j) - \boldsymbol{q}(t_{j-1}) - \tau\mathbb{Q}\widehat{\boldsymbol{q}}(t_j \mid t_{j-1}) \\
&= \boldsymbol{q}(t_j) - \boldsymbol{q}(t_{j-1}) - \tau\mathbb{Q}\boldsymbol{q}(t_{j-1}) - \tau^2\mathbb{Q}^2\widehat{\boldsymbol{q}}(t_j \mid t_{j-1}) \\
&= \boldsymbol{q}(t_j) - \boldsymbol{q}(t_{j-1}) - \tau\mathbb{Q}\boldsymbol{q}(t_{j-1}) - \tau^2\mathbb{Q}^2\boldsymbol{q}(t_{j-1}) + \mathcal{O}(\tau^3),
\end{aligned} \tag{B.9}$$

where we have used Eq. (B.8) twice. A Taylor series expansion of $\boldsymbol{q}(t_{j-1}+\tau)$ around t_{j-1} gives

$$\begin{aligned}
\boldsymbol{q}(t_j) &= \boldsymbol{q}(t_{j-1}+\tau) \\
&= \boldsymbol{q}(t_{j-1}) + \tau\frac{d\boldsymbol{q}(t_{j-1})}{dt} + \frac{1}{2}\,\tau^2\frac{d^2\boldsymbol{q}(t_{j-1})}{dt^2} + \mathcal{O}(\tau^3) \\
&= \boldsymbol{q}(t_{j-1}) + \tau\mathbb{Q}\boldsymbol{q}(t_{j-1}) + \frac{1}{2}\,\tau^2\mathbb{Q}^2\boldsymbol{q}(t_{j-1}) + \mathcal{O}(\tau^3),
\end{aligned} \tag{B.10}$$

by virtue of Eq. (3.2.3), which, together with Eq. (B.9), results in

$$\boldsymbol{q}(t_j) = \widehat{\boldsymbol{q}}_\tau(t_j \mid t_{j-1}) - \frac{1}{2}\,\tau^2\mathbb{Q}^2\boldsymbol{q}(t_{j-1}) + \mathcal{O}(\tau^3), \tag{B.11}$$

where we now use $\widehat{\boldsymbol{q}}_\tau(t_j \mid t_{j-1})$ to denote the fact that the approximate solution $\widehat{\boldsymbol{q}}(t_j \mid t_{j-1})$ is obtained with step-size τ.

APPENDIX B

Let us now denote by $\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})$ the approximate solution obtained by the IE method at time t_j when the step-size is $\tau/2$. Note that $\widehat{\boldsymbol{q}}_{\tau/2}(t_{j-1}+\tau/2 \mid t_{j-1}) = (\mathbb{I}-\tau\mathbb{Q}/2)^{-1}\boldsymbol{q}(t_{j-1})$ and $\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) = (\mathbb{I}-\tau\mathbb{Q}/2)^{-1}\widehat{\boldsymbol{q}}_{\tau/2}(t_{j-1}+\tau/2 \mid t_{j-1})$, by virtue of Eq. (3.2.4). Therefore, $\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) = (\mathbb{I}-\tau\mathbb{Q}/2)^{-2}\boldsymbol{q}(t_{j-1})$, or [compare with Eq. (B.8)]

$$\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) = \boldsymbol{q}(t_{j-1}) + \tau\mathbb{Q}\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) - \frac{\tau^2}{4}\mathbb{Q}^2\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}). \tag{B.12}$$

We now have that

$$\begin{aligned}
\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) &= \boldsymbol{q}(t_j) - \boldsymbol{q}(t_{j-1}) - \tau\mathbb{Q}\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) + \frac{\tau^2}{4}\mathbb{Q}^2\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) \\
&= \boldsymbol{q}(t_j) - \boldsymbol{q}(t_{j-1}) - \tau\mathbb{Q}\boldsymbol{q}(t_{j-1}) + \frac{\tau^2}{4}\mathbb{Q}^2\boldsymbol{q}(t_{j-1}) - \tau^2\mathbb{Q}^2\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) + \mathcal{O}(\tau^3) \\
&= \boldsymbol{q}(t_j) - \boldsymbol{q}(t_{j-1}) - \tau\mathbb{Q}\boldsymbol{q}(t_{j-1}) - \frac{3}{4}\,\tau^2\mathbb{Q}^2\boldsymbol{q}(t_{j-1}) + \mathcal{O}(\tau^3),
\end{aligned} \tag{B.13}$$

where we have used Eq. (B.12) twice. From the Taylor series expansion Eq. (B.10) and Eq. (B.13), we finally obtain

$$\boldsymbol{q}(t_j) = \widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) - \frac{1}{4}\,\tau^2\mathbb{Q}^2\boldsymbol{q}(t_{j-1}) + \mathcal{O}(\tau^3). \tag{B.14}$$

Now, from Eq. (B.11) and Eq. (B.14), we have

$$\boldsymbol{q}(t_j) = 2\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) - \widehat{\boldsymbol{q}}_{\tau}(t_j \mid t_{j-1}) + \mathcal{O}(\tau^3). \tag{B.15}$$

This result shows that

$$\widehat{\boldsymbol{q}}_{*}(t_j \mid t_{j-1}) := 2\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) - \widehat{\boldsymbol{q}}_{\tau}(t_j \mid t_{j-1})$$

may produce a better approximation to $\boldsymbol{q}(t_j)$ than either $\widehat{\boldsymbol{q}}_{\tau}(t_j \mid t_{j-1})$ or $\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})$, since it results in a third-order approximation (in terms of the local error) of $\boldsymbol{q}(t_j)$, as compared to $\widehat{\boldsymbol{q}}_{\tau}(t_j \mid t_{j-1})$ or $\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})$ which result in second-order approximations.

We can also use $\widehat{\boldsymbol{q}}_{\tau}(t_j \mid t_{j-1})$ and $\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})$ to determine an appropriate step-size τ_* that guarantees a local error within a pre-specified tolerance TOL. Indeed, if we define the local error $\mathrm{ERR} := \|\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})\|_1$, then from Eq. (B.15), we approximately have that

$$\mathrm{ERR} = \|\widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) - \widehat{\boldsymbol{q}}_{\tau}(t_j \mid t_{j-1})\|_1, \tag{B.16}$$

which provides a way to calculate the error for a sufficiently small step-size τ. If now ERR $\neq$ TOL, then we need to change the step-size to a new value τ_*, such that $\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau_*/2}(t_j \mid t_{j-1}) = \boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})$, which will imply that TOL $:= \|\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau_*/2}(t_j \mid t_{j-1})\|_1 = \|\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})\|_1 =$ ERR. From Eq. (B.14), we have that

$$\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1}) \simeq -\frac{1}{4}\,\tau^2\mathbb{Q}^2\boldsymbol{q}(t_{j-1})$$

$$\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau_*/2}(t_j \mid t_{j-1}) \simeq -\frac{1}{4}\,\tau_*^2\mathbb{Q}^2\boldsymbol{q}(t_{j-1}),$$

from which we obtain

$$\frac{\tau_*^2}{\tau^2} \simeq \frac{\|\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau_*/2}(t_j \mid t_{j-1})\|_1}{\|\boldsymbol{q}(t_j) - \widehat{\boldsymbol{q}}_{\tau/2}(t_j \mid t_{j-1})\|_1} = \frac{\text{TOL}}{\text{ERR}}\,.$$

As a consequence, the desired step-size value will approximately be given by

$$\tau^* = \tau\sqrt{\frac{\text{TOL}}{\text{ERR}}}\,. \tag{B.17}$$

Expokit parameter values

The required TOL value (used to determine a desired error tolerance for the KSA method and the RIE method with variable step-size) was set to 1×10^{-3}. We obtained the Expokit solution by using a Krylov subspace approximation with dimension $L_0 = 65$. This value was obtained by starting with the default value of $L_0 = 30$ and successively increasing it by 5 until the resulting Expokit error estimate was less than TOL $= 1 \times 10^{-3}$. The reported L2 errors were calculated using a solution obtained by a computationally more expensive Expokit run with $L_0 = 70$ and TOL $= 1 \times 10^{-4}$, which we consider it to be the 'true' solution. This is based on the premise that Expokit will produce the true solution for sufficiently large L_0 and small TOL.

Comparison of IE to an augmented state space

Here we discuss how the previously published results in [139] apply to the present framework of N species X_n, $n \in \mathcal{N}$, governed by the reactions of Eq. (2.1.1) with corresponding propensities

$\pi_m(\boldsymbol{x})$, $m \in \mathcal{M}$. The idea is to define an ordering on an augmented state space such that a reaction can only take the state from a lesser value to a greater value with respect to this ordering. Then by arranging the states from this augmented state space in the probability vector $\boldsymbol{p}_a(t)$ in increasing order, one ensures the resulting generator matrix $\mathbb{P}_a$ will be lower triangular.

To define this ordering, we augment the state space with an additional "counting" species, X_{N+1}, whose value monotonically increases in a way that guarantees the resulting augmented population (AP) process can always be well ordered. The new system has species $\widetilde{X}_1, \ldots, \widetilde{X}_N, \widetilde{X}_{N+1}$ which have a well ordered state space according to the rule $\widetilde{\boldsymbol{x}} < \widetilde{\boldsymbol{x}}'$ whenever

1) $\sum_{n=1}^{N+1} \widetilde{x}_n < \sum_{n=1}^{N+1} \widetilde{x}'_n$, or

2) $\sum_{n=1}^{N+1} \widetilde{x}_n = \sum_{n=1}^{N+1} \widetilde{x}'_n$, and $\widetilde{x}_{N+1} < \widetilde{x}'_{N+1}$, or

3) $\sum_{n=1}^{N+1} \widetilde{x}_n = \sum_{n=1}^{N+1} \widetilde{x}'_n$, and $\widetilde{x}_{N+1} = \widetilde{x}'_{N+1}$, and $\widetilde{x}_N < \widetilde{x}'_N$, or

$\vdots$

N+1) $\sum_{n=1}^{N+1} \widetilde{x}_n = \sum_{n=1}^{N+1} \widetilde{x}'_n$, and $\widetilde{x}_{N+1} = \widetilde{x}'_{N+1}$,..., and $\widetilde{x}_2 < \widetilde{x}'_2$.

These new species are governed by the following M reactions

$$\sum_{n=1}^{N} \nu_{nm} \widetilde{X}_n \to \nu'_{N+1,m} \widetilde{X}_{N+1} + \sum_{n=1}^{N} \nu'_{nm} \widetilde{X}_n, \quad m \in \mathcal{M},$$

with corresponding propensities $\widetilde{\pi}_m(\widetilde{x}_1, \ldots, \widetilde{x}_{N+1}) = \pi_m(\widetilde{x}_1, \ldots, \widetilde{x}_N), m \in \mathcal{M}$ that are equal to the original propensities, whereas $\nu'_{N+1,m}$ is chosen such that the reactions always move the state from a lower value to a higher value with respect to the ordering. Note that this means the dynamics of $\widetilde{Z}_1, \ldots, \widetilde{Z}_M$ remain identical to $Z_1, \ldots, Z_M$, and thus the marginal distribution of $\widetilde{X}_1, \ldots, \widetilde{X}_N$ remains equivalent to the original joint distribution on $X_1, \ldots, X_N$.

In order to chose $\nu'_{N+1,m}$, note that there are three types of reactions in the original system (2.1.1): **(i)** those such that $\sum_{n=1}^{N} s_{nm} > 0$, **(ii)** those such that $\sum_{n=1}^{N} s_{nm} = 0$, and **(iii)** those such that $\sum_{n=1}^{N} s_{nm} < 0$. If reaction m is of type **(i)**, then $\nu'_{N+1,m}$ can be set equal to zero

and this reaction will result in the state variable increasing according to rule **1)**. If reaction m is of type **(ii)**, then $\nu'_{N+1,m}$ can be set equal to one and this reaction will result in the state variable increasing according to rule **2)**. If reaction m is of type **(iii)**, then $\nu'_{N+1,m}$ can be set equal to $-\sum_{n=1}^{N} s_{nm}$ and this reaction will result in the state variable increasing according to rule **2)**. These conclusions lead us to choose the stoichiometry of the counting variable

$$\nu'_{N+1,m} := \left[\sum_{n=1}^{N} s_{nm} < 0\right]\left(-\sum_{n=1}^{N} s_{nm}\right) + \left[\sum_{n=1}^{N} s_{nm} = 0\right],$$

where $[\cdot]$ is the Iverson bracket that takes value one when its argument is true and zero otherwise.

As noted earlier, the marginal distribution of $\widetilde{X}_1, \ldots, \widetilde{X}_N$ remains unchanged from the distribution of the original system (2.1.1). Thus, after the joint probabilities over $\widetilde{X}_1, \ldots, \widetilde{X}_{N+1}$ are computed in this augmented state space it is necessary to marginalize the distribution over the counting variable $\widetilde{X}_{N+1}$ to find the joint probability distribution of the original system.

Example: Let us consider a simple example along these lines. Suppose the reactions that govern X_1, X_2 are given by

$$X_1 \rightarrow X_1 + X_2 \tag{B.18}$$

$$X_2 \rightarrow X_1 \tag{B.19}$$

$$2X_1 \rightarrow \emptyset, \tag{B.20}$$

with propensities given by $\pi_1(\boldsymbol{x}) = k_1 x_1, \pi_2(\boldsymbol{x}) = k_2 x_2, \pi_3(\boldsymbol{x}) = k_3 x_1^2$. Now we wish to augment this state space with a counting variable X_3 to ensure that the resulting system is triangular. Here reaction (B.18) has $\sum_{n=1}^{2} s_{n1} = 1$ which means that it always increases the state with respect to the ordering according to rule **1)**, so the count variable need not increase when this reaction occurs. Thus $\nu_{3,1} = 0$. On the other hand reaction (B.19) has $\sum_{n=1}^{2} s_{n2} = 0$ and thus we set $\nu_{3,2} = 1$ to ensure that this reaction results in the state increasing its order according to rule **1)**. Finally, reaction (B.20) has $\sum_{n=1}^{2} s_{n2} = -2$ and thus we set $\nu_{3,2} = 2$ to ensure that this reaction results in the state increasing its order according to rule **2)**. This results in the augmented system being given

by

$$\begin{aligned} \widetilde{X}_1 &\rightarrow \widetilde{X}_1 + \widetilde{X}_2 \\ \widetilde{X}_2 &\rightarrow \widetilde{X}_1 + \widetilde{X}_3 \\ 2\widetilde{X}_1 &\rightarrow 2\widetilde{X}_3, \end{aligned}$$

with propensities given by $\widetilde{\pi}_1(\widetilde{\boldsymbol{x}}) = k_1\widetilde{x}_1, \widetilde{\pi}_2(\widetilde{\boldsymbol{x}}) = k_2\widetilde{x}_2, \widetilde{\pi}_3(\widetilde{\boldsymbol{x}}) = k_3\widetilde{x}_1^2$. It is easily verified that for any chosen value of $\widetilde{\boldsymbol{x}}$, each reaction will the system move to a new state $\widetilde{\boldsymbol{x}}'$ such that $\widetilde{\boldsymbol{x}} < \widetilde{\boldsymbol{x}}'$ according to the aforementioned ordering. ♠

The aforementioned method is intimately related to the DA process. Intuitively, the reason this augmentation method works is that it leaves the DA dynamics undisturbed, and then simply adds a new row to the stoichiometry matrix (which maps the DA process to the population process). Thus the DA process is more fundamental than the AP process, and this new row of the stoichiometry matrix is designed to "capture" the monotonicity of the DA process in the augmented state space. The major disadvantage of the DA process is the fact that it may become unbounded in some systems, but unfortunately, the AP process will be unbounded whenever the DA process is unbounded. The contrapositive of this statement is that if the AP process is finite, then so too is the DA process. Thus, the AP process provides no obvious advantage over the DA process, but the simple and fundamental nature of the DA process (in particular the lexicographical ordering which has the advantage of arising from nested for-loops in computer algorithms) leads us to conclude that the DA process is more appropriate for use with the IE method.

Proof: We will prove the statement "If the DA process is unbounded, so too is the augmented state space process." It is clear that Z_m being unbounded for any reaction which has $\nu_{N+1,m} > 0$ will cause the AP process to become unbounded since the counting variable X_{N+1} can never be decremented. Thus, we must only show that Z_m being infinite over at least one of the reactions with $\nu_{N+1,m} = 0$ results in the AP process being unbounded.

Suppose by way of contradiction that there exists a Z_m which is unbounded for a reaction

with $\nu_{N+1,m} = 0$ and that the AP process remains bounded. If $\nu_{N+1,m} = 0$, then $\sum_{n=1}^{N} s_{nm} > 0$. Thus, the unbounded reaction m increments the value $\sum_{n=1}^{N} X_n$ an unbounded number of times. Since the AP process is assumed to be bounded, this means that at least one reaction m' must decrement the value $\sum_{n=1}^{N} X_n$ an unbounded number of times. As a consequence, there exists an unbounded reaction $Z_{m'}$ that has $\sum_{n=1}^{N} s_{nm'} < 0$ or, equivalently, $\nu_{N+1,m'} = -\sum_{n=1}^{N} s_{nm'} > 0$, which implies that X_{N+1} is unbounded. This contradicts our initial assumptions and concludes the proof. ♠

Severo provided a recursive procedure for computing the solution of a triangular system of differential equations [140] that could be used instead of the IE computations considered here. In particular, Severo showed that the solution to the equation $\frac{d\boldsymbol{q}(t)}{dt} = \mathbb{Q}\boldsymbol{q}(t)$ is given by $\boldsymbol{q}(t) = \mathbb{C}(t)\mathbf{d}(t)$. Here $\mathbb{C}(t)$ is a lower triangular $K \times K$ matrix, whose elements are $\binom{K+1}{2}$ polynomials in t with coefficients that can be computed recursively, while $\mathbf{d}(t)$ is found by evaluating $d_k(t) = \exp(q_{kk}t)$ for $k = 1, \ldots, K$ [141]. In the specific case of the SIR model with a special initial condition, a simpler recursive procedure than Severo's general recursion has been found for computing the coefficients of the polynomials in $\mathbb{C}(t)$ [244]. Even with the $\mathbb{C}(t)$ matrix and the $\mathbf{d}(t)$ vector computed off-line, evaluating their product is significantly more costly than performing the IE iterations because the matrix $\mathbb{C}(t)$ looses much of the sparse structure present in the $\mathbb{Q}$ matrix. In particular, a sufficient condition for the $\mathbb{C}(t)$ matrix to have an element equal to zero, e.g., $c_{kk'}(t) = 0$, is for the states $\boldsymbol{z}_{k'}$ and $\boldsymbol{z}_k$ to be non-communicating in the Markov chain [141]. Comparatively, a sufficient condition for the $\mathbb{Q}$ matrix to have an element equal to zero, e.g., $a_{kk'} = 0$, is that $\boldsymbol{z}_{k'}$ is not adjacent to $\boldsymbol{z}_k$ in the Markov chain. Thus, a single matrix vector multiplication $\mathbb{C}(t)\mathbf{d}(t)$ will in general require $\mathcal{O}(K^2)$ operations, as compared to the $\mathcal{O}(K)$ operations of the entire IE method.

This massive added overhead of recursively computing the coefficients of $\mathbb{C}(t)$, evaluating the $\binom{K+1}{2}$ polynomials in $\mathbb{C}(t)$ for each time point of interest, and then evaluating the K exponentials at each time point of interest to find $\mathbf{d}(t)$, only add to the advantage of the IE method. Furthermore, the storage requirements of a particular $\mathbb{C}(t)$ and the coefficients of its polynomials rapidly exceed

the requirements of storing the $\mathbb{Q}$ matrix. Severo's procedure is thus impractical for the large values of K considered in this thesis. Its main utility remains in finding analytical solutions in systems with very small K or proving a particular functional form of a solution to a specific system. This procedure has been used in the literature [139–141,244] for small values of K much before the current computing power became available.

Appendix C

Estimating the true critical value in KS testing

Given the null hypothesis, the solution of the master equation is assumed to be given by the analytical approximation. Keeping this in mind, for each $k = 1, 2, \ldots, K$, we can independently draw L samples $\{\widehat{x}_1^{(k)}(t), \widehat{x}_2^{(k)}(t), \ldots, \widehat{x}_L^{(k)}(t)\}$ from the analytical approximation $\widehat{P}(x;t)$ and use these samples to compute K independent samples

$$G^{(k)}(x;t) = \frac{1}{L}\sum_{l=1}^{L}[\widehat{x}_l^{(k)}(t) \leq x], \quad k = 1, 2, \ldots, K, \tag{C.1}$$

of the empirical CDF and, subsequently, K independent samples

$$S^{(k)}(t) = \max_x \left|G^{(k)}(x;t) - \widehat{F}(x;t)\right|, \quad k = 1, 2, \ldots, K, \tag{C.2}$$

of the KS statistic. We can then order the latter samples in an increasing order $S^{[1]}(t) \leq S^{[2]}(t) \leq \cdots \leq S^{[K]}(t)$, and approximate the critical value by (see pp. 219 & 221 in Ref. [149])

$$s_0(\alpha;t) \simeq S^{[(K+1)(1-\alpha)+1]}(t), \tag{C.3}$$

where K is taken to be a value that ensures $(K+1)(1-\alpha)+1$ is an integer.

Choosing the significance level and sample size in Scenario 3

We here discuss a method for determining appropriate values for the significance level α and the sample size L of the proposed KS hypothesis testing procedure so that the probability of

the Type I error and the power of the test are contained within allowable values. We focus our discussion here on testing the validity of the LNA method.

The *power* of a test against an alternative distribution $Q \in \mathcal{P}_A$ is the probability of rejecting the null hypothesis when the true distribution is given by Q. The test may have a different power against different alternative distributions. One may think of the power of a test against Q as quantifying how easily the test can tell that this distribution is not the null distribution. An alternative distribution with low power looks like the null distribution to the test, whereas an alternative distribution with higher power can be distinguished easier by the test from the null distribution. We will denote the power of the proposed KS test against Q by $\beta_{\alpha,L}[Q]$, where the subscripts remind us that this quantity depends on the significance level α and the sample size L.

Ideally, whenever the LNA method provides a poor approximation to an alternative distribution Q, we would like $\beta_{\alpha,L}[Q] = 1$, so that the test would always reject the validity of this approximation method. However, we do not want the power to be always one, since there might be some alternative distribution for which the LNA method provides a good enough approximation. In such a case, we would like $\beta_{\alpha,L}[Q] = 0$ (i.e., we would like not to reject the validity of the LNA method with probability one).

Let us consider the alternative (mixture) distributions given by Eq. (4.3.16) and assume that the system is initialized within the basin of attraction of the stable fixed point $\Omega\mu^*$. For sufficiently large r, Q_r will be almost identical to the Gaussian distribution $\widehat{P}^*$ predicted by the LNA method. In this case, we would like the power to be zero. For sufficiently small r however the LNA method produces a poor approximation, since most probability mass of Q_r will now be concentrated at the second stable fixed point $\Omega\mu^{**}$ and we would like the power to be one in this case. Of course, it is not possible to design a hypothesis test with such an ideal power. For this reason, we must focus our attention on designing a test based on more realistic criteria.

To proceed, we will first examine the behavior of the power $\beta_{\alpha,L}[Q_r]$ as a function of L, r, and α. It turns out that the KS test satisfies $\lim_{L\to\infty} \beta_{\alpha,L}[Q] = 1$, for any $Q \in \mathcal{P}_A$ [148].

We thus expect $\beta_{\alpha,L}[Q_r]$ to be a monotonically increasing function of L, since the test obtains more discriminatory power with each additional sample. We also expect $\beta_{\alpha,L}[Q_r]$ to monotonically increase as $r \to 0$, since Q_r becomes increasingly different than $\widehat{P}^*$ for smaller values of r. As a matter of fact, we expect that $\lim_{r\to 0} \beta_{\alpha,L}[Q_r] = 1$ for large enough L. Note finally that $\lim_{r\to 1} \beta_{\alpha,L}[Q_r] = \alpha$, since $r = 1$ corresponds to the null hypothesis which will be rejected with probability α. Finally, $\beta_{\alpha_1,L}[Q_r] \leq \beta_{\alpha_2,L}[Q_r]$, when $\alpha_1 < \alpha_2$, since increasing the significance level increases the probability of rejecting the null hypothesis and thus increases the power.

We can verify the previously described behavior by estimating the power $\beta_{\alpha,L}[Q_r]$ of the KS test, for any values of r, α and L, using Monte Carlo sampling. Since the power $\beta_{\alpha,L}[Q_r]$ is the probability that the test rejects the null hypothesis when the true distribution is Q_r, we can perform the level-α KS hypothesis test numerous times using L samples drawn independently from Q_r and estimate $\beta_{\alpha,L}[Q_r]$ as the fraction of times this procedure rejects the null hypothesis. We depict the results of this estimation in Fig. C.1.

Now that we have characterized the behavior of the power, we can partition the values of r into three regions, a *rejection*, an *acceptance* and an *indifference* region, and specify the desired level of power within each region. The rejection region $[0, r_0]$ consists of all r values for which we would like the test to reject the null hypothesis (i.e., the region in which the LNA method produces a poor approximation). Although the power of the test should ideally be one in this region, in practice, we should specify a minimum allowable power $0 \ll b_0 < 1$, in which case we will have $b_0 \leq \beta_{\alpha,L}[Q_r] \leq 1$, for all $0 \leq r \leq r_0$. Likewise, the acceptance region $[r_1, 1]$ consists of all r values for which we want not to reject the null hypothesis (i.e., the region in which the LNA method is considered to be valid). Ideally, the test should have zero power in this region. In practice however we must specify a maximum allowable power $0 < b_1 \ll 1$, in which case we will have $0 \leq \beta_{\alpha,L}[Q_r] \leq b_1$, for all $r_1 \leq r \leq 1$. In the remaining region (r_0, r_1), known as the indifference region, the power of the test transitions from high to low values. We are not much interested in this region because the LNA method produces neither a good nor a bad approximation to the true probability distribution.

APPENDIX C

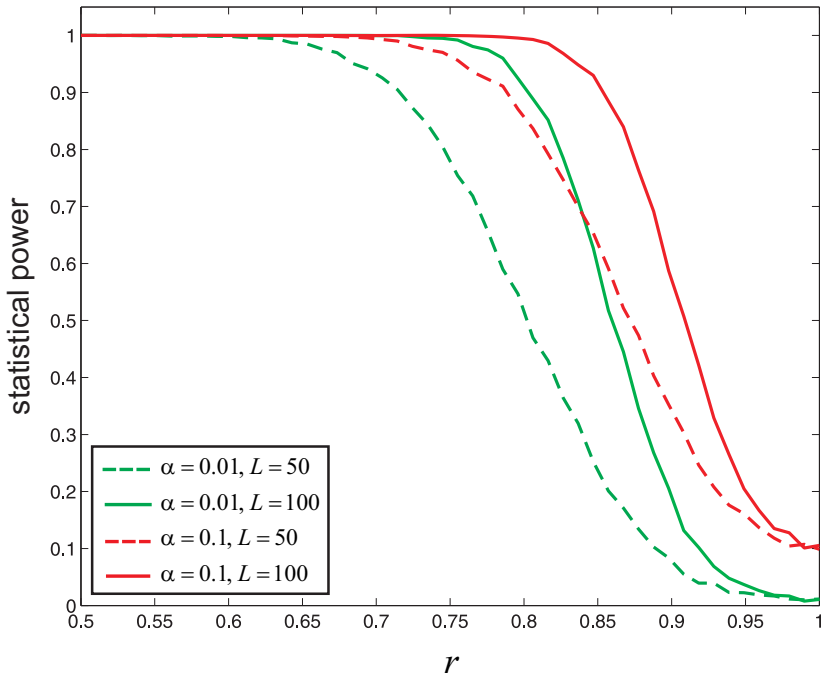

Figure C.1: The power $\beta_{\alpha,L}[Q_r]$ of the KS test estimated by Monte Carlo sampling, plotted as a function of r and for various values of α and L. The distribution $\widehat{P}^*$ used here is a normal distribution with mean 0.5 and standard deviation 0.05, whereas, the distribution $\widehat{P}^{**}$ is a normal distribution with mean 0.8 and standard deviation 0.02; see Eq. (4.3.16). Each value of $\beta_{\alpha,L}[Q_r]$ was computed using 4,000 Monte Carlo samples.

It is clear from Fig. C.1 that we can design a test that meets the previous criteria by choosing a small enough significance level α, to ensure that the power is small in the acceptance region, and a large enough sample size L, to ensure that the power is large in the rejection region. We can do this by the following very simple iterative procedure:

1. We start with an initial value for L and set the value of α to be the maximum allowable probability of the Type I error.
2. While keeping the value of L fixed, we decrease the value of α until the estimated power satisfies the desired criterion (i.e., it is smaller than b_1) in the acceptance region.
3. While keeping the value of α fixed, we increase the value of L until the estimated power satisfies the desired criterion (i.e., it is larger than b_0) in the rejection region.
4. If the estimated power in step 3 also satisfies the criterion in the acceptance region, we stop; otherwise we go to step 2 and repeat the procedure.

Appendix D

Markovianity of fractional activity process in LMNs

Because the activity process $\{\boldsymbol{X}(t), t \geq 0\}$ is a Markov process, only a single transition from $\boldsymbol{X}(t)$ to $\boldsymbol{X}(t) + \boldsymbol{e}_{n^*}$ or $\boldsymbol{X}(t) - \boldsymbol{e}_{n^*}$, for some $n^* \in \mathcal{N}$, can occur within the infinitesimally small time interval $[t, t + dt)$, where $\boldsymbol{e}_{n^*}$ is the n^*-th column of the $N \times N$ identity matrix. Since

$$Y_k(t) := \frac{1}{N_k} \sum_{n \in \mathcal{N}_k} X_n(t), \quad \text{for every } k \in \mathcal{K}, \tag{D.1}$$

where $\mathcal{K} = \{1, 2, \ldots, K\}$, with K being the number of homogeneous sub-populations $\mathcal{N}_k$, and $N_k = |\mathcal{N}_k|$ (i.e., the cardinality of $\mathcal{N}_k$), this also means that the fractional activity process $\{\boldsymbol{Y}(t), t \geq 0\}$ can transition only once within $[t, t + dt)$ from $\boldsymbol{Y}(t)$ to $\boldsymbol{Y}(t) + \widetilde{\boldsymbol{e}}_k$ or $\boldsymbol{Y}(t) - \widetilde{\boldsymbol{e}}_k$. Here, $\widetilde{\boldsymbol{e}}_k$ is the k-th column of the $K \times K$ identity matrix multiplied by N_k^{-1} and $k \in \mathcal{K}$ is such that $n^* \in \mathcal{N}_k$. As a matter of fact, the transition probabilities are given by

$$\begin{aligned}
&\Pr\left[\boldsymbol{Y}(t) \to \boldsymbol{Y}(t) + \widetilde{\boldsymbol{e}}_k \text{ within } [t, t + dt) \mid \boldsymbol{X}(t) = \boldsymbol{x}\right] \\
&\quad = \sum_{n \in \mathcal{N}_k} p_n^+(\boldsymbol{x}) dt \\
&\quad = \sum_{n \in \mathcal{N}_k} (1 - x_n)\left[\ell_n^+ + f_n(r_n(\boldsymbol{x}))\right] dt \\
&\quad = \sum_{n \in \mathcal{N}_k} (1 - x_n)\left[\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{y}))\right] dt \\
&\quad = \Big(N_k - \sum_{n \in \mathcal{N}_k} x_n\Big)\left[\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{y}))\right] dt \\
&\quad = N_k(1 - y_k)\left[\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{y}))\right] dt,
\end{aligned} \tag{D.2}$$

where $p_n^+(\boldsymbol{x})$ is the propensity function of activation of the n-th node of the LMN and the k-th element y_k of $\boldsymbol{y}$ is given by $N_k^{-1}\sum_{n\in\mathcal{N}_k} x_n$. Moreover, ϕ_k and λ_k^+ are such that, for every $k \in \mathcal{K}$, $f_n = \phi_k$ and $\ell_n^+ = \lambda_k^+$, for all $n \in \mathcal{N}_k$, and we have ignored terms that go to zero faster than dt. Finally, we have used the assumption that there exists a function ρ_k such that $r_n(\boldsymbol{x}) = \rho_k(\boldsymbol{y})$, for every $n \in \mathcal{N}_k$. Likewise, we have that

$$\begin{aligned}
&\Pr\left[\boldsymbol{Y}(t) \to \boldsymbol{Y}(t) - \widetilde{\boldsymbol{e}}_k \text{ within } [t, t+dt) \mid \boldsymbol{X}(t) = \boldsymbol{x}\right] \\
&\quad = \sum_{n\in\mathcal{N}_k} p_n^-(\boldsymbol{x})dt \\
&\quad = \sum_{n\in\mathcal{N}_k} x_n[\ell_n^- + g_n(r_n(\boldsymbol{x}))]dt \\
&\quad = \sum_{n\in\mathcal{N}_k} x_n[\lambda_k^- + \gamma_k(\rho_k(\boldsymbol{y}))]dt \\
&\quad = N_k y_k[\lambda_k^- + \gamma_k(\rho_k(\boldsymbol{y}))]dt, \qquad \text{(D.3)}
\end{aligned}$$

where $p_n^-(\boldsymbol{x})$ is the propensity function of inactivation of the n-th node of the LMN, whereas γ_k and λ_k^- are such that, for every $k \in \mathcal{K}$, $g_n = \gamma_k$ and $\ell_n^- = \lambda_k^-$, for all $n \in \mathcal{N}_k$.

The previous discussion shows that the fractional activity process $\{\boldsymbol{Y}(t), t \geq 0\}$ is Markovian with propensity functions given by

$$\pi_k^+(\boldsymbol{y}) \;=\; N_k(1-y_k)[\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{y}))] \qquad \text{(D.4)}$$

$$\pi_k^-(\boldsymbol{y}) \;=\; N_k y_k[\lambda_k^- + \gamma_k(\rho_k(\boldsymbol{y}))]. \qquad \text{(D.5)}$$

As a consequence, the probability distribution $P(\boldsymbol{y}; t)$ satisfies the master equation 5.2.13. Finally, it is important to note that, when the net input to a node n in the LMN is given by $r_n(\boldsymbol{x}) = h_n + \boldsymbol{a}_n^T\boldsymbol{x}$, we can find a function ρ_k such that $r_n(\boldsymbol{x}) = \rho_k(\boldsymbol{y})$, for every $n \in \mathcal{N}_k$. Indeed, if $h_n = \eta_k$, for every

$n \in \mathcal{N}_k$, and $a_{nn'} = w_{kk'}/N_{k'}$, for every $n \in \mathcal{N}_k$, $n' \in \mathcal{N}_{k'}$, then, for every $n \in \mathcal{N}_k$, we have that

$$\begin{aligned} r_n(\boldsymbol{x}) &= h_n + \sum_{n' \in \mathcal{N}} a_{nn'} x_{n'} \\ &= h_n + \sum_{k' \in \mathcal{K}} \sum_{n' \in \mathcal{N}_{k'}} a_{nn'} x_{n'} \\ &= \eta_k + \sum_{k' \in \mathcal{K}} \sum_{n' \in \mathcal{N}_{k'}} \frac{w_{kk'}}{N_{k'}} \, x_{n'} \\ &= \eta_k + \sum_{k' \in \mathcal{K}} w_{kk'} \, \frac{1}{N_{k'}} \sum_{n' \in \mathcal{N}_{k'}} x_{n'} \\ &= \eta_k + \sum_{k' \in \mathcal{K}} w_{kk'} y_{k'}, \end{aligned} \tag{D.6}$$

which shows that $\rho_k(\boldsymbol{y}) = \eta_k + \sum_{k' \in \mathcal{K}} w_{kk'} y_{k'}$.

Linear Noise Approximation in LMNs

By following [1] (see also [19]), we define the shift operators Σ_k^- and Σ_k^+ by

$$\Sigma_k^- \varphi(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, \ldots, y_K) := \varphi(y_1, \ldots, y_{k-1}, y_k - N_k^{-1}, y_{k+1}, \ldots, y_K) \tag{D.7}$$

$$\Sigma_k^+ \varphi(y_1, \ldots, y_{k-1}, y_k, y_{k+1}, \ldots, y_K) := \varphi(y_1, \ldots, y_{k-1}, y_k + N_k^{-1}, y_{k+1}, \ldots, y_K), \tag{D.8}$$

for $k \in \mathcal{K}$ and any function $\varphi(\boldsymbol{y})$. By using a Taylor series expansion, we have that

$$\Sigma_k^- \varphi(\boldsymbol{y}) = \varphi(\boldsymbol{y}) - N_k^{-1} \frac{\partial \varphi(\boldsymbol{y})}{\partial y_k} + \frac{N_k^{-2}}{2} \frac{\partial^2 \varphi(\boldsymbol{y})}{\partial y_k^2} - \frac{N_k^{-3}}{3!} \frac{\partial^3 \varphi(\boldsymbol{y})}{\partial y_k^3} + \cdots \tag{D.9}$$

$$\Sigma_k^+ \varphi(\boldsymbol{y}) = \varphi(\boldsymbol{y}) + N_k^{-1} \frac{\partial \varphi(\boldsymbol{y})}{\partial y_k} + \frac{N_k^{-2}}{2} \frac{\partial^2 \varphi(\boldsymbol{y})}{\partial y_k^2} + \frac{N_k^{-3}}{3!} \frac{\partial^3 \varphi(\boldsymbol{y})}{\partial y_k^3} + \cdots \tag{D.10}$$

In this case, we can write the master equation (5.2.13) as

$$\begin{aligned} \frac{\partial P(\boldsymbol{y};t)}{\partial t} &= \sum_{k \in \mathcal{K}} \Big\{ \left(\Sigma_k^- - 1\right) \pi_k^+(\boldsymbol{y}) P(\boldsymbol{y};t) + \left(\Sigma_k^+ - 1\right) \pi_k^-(\boldsymbol{y}) P(\boldsymbol{y};t) \Big\} \\ &= \sum_{k \in \mathcal{K}} \Big\{ \Big(- N_k^{-1} \frac{\partial}{\partial y_k} + \frac{N_k^{-2}}{2} \frac{\partial^2}{\partial y_k^2} - \frac{N_k^{-3}}{3!} \frac{\partial^3}{\partial y_k^3} + \cdots \Big) \times \pi_k^+(\boldsymbol{y}) P(\boldsymbol{y};t) \\ &\qquad + \Big(N_k^{-1} \frac{\partial}{\partial y_k} + \frac{N_k^{-2}}{2} \frac{\partial^2}{\partial y_k^2} + \frac{N_k^{-3}}{3!} \frac{\partial^3}{\partial y_k^3} + \cdots \Big) \times \pi_k^-(\boldsymbol{y}) P(\boldsymbol{y};t) \Big\}. \end{aligned} \tag{D.11}$$

APPENDIX D

If we now define the drift and diffusion functions by

$$A_k(\boldsymbol{y}) := \frac{\pi_k^+(\boldsymbol{y}) - \pi_k^-(\boldsymbol{y})}{N_k} \tag{D.12}$$

$$D_k(\boldsymbol{y}) := \frac{\pi_k^+(\boldsymbol{y}) + \pi_k^-(\boldsymbol{y})}{N_k}, \tag{D.13}$$

respectively, then

$$\frac{\partial P(\boldsymbol{y};t)}{\partial t} = -\sum_{k\in\mathcal{K}} \Big\{ \frac{\partial[A_k(\boldsymbol{y})P(\boldsymbol{y};t)]}{\partial y_k} - \frac{N_k^{-1}}{2}\frac{\partial^2[D_k(\boldsymbol{y})P(\boldsymbol{y};t)]}{\partial y_k^2} + \frac{N_k^{-2}}{3!}\frac{\partial^3[A_k(\boldsymbol{y})P(\boldsymbol{y};t)]}{\partial y_k^3} + \cdots \Big\}. \tag{D.14}$$

This equation, together with the *ansatz*

$$Y_k(t) = \mu_k(t) + \frac{1}{\sqrt{N_k}}\, W_k(t), \quad t > 0, \quad k \in \mathcal{K}, \tag{D.15}$$

yields

$$\frac{\partial P(\boldsymbol{y};t)}{\partial t} = -\sum_{k\in\mathcal{K}} \Big\{ N_k^{1/2}\frac{\partial[A_k(\boldsymbol{y})P(\boldsymbol{y};t)]}{\partial w_k} - \frac{1}{2}\frac{\partial^2[D_k(\boldsymbol{y})P(\boldsymbol{y};t)]}{\partial w_k^2} + \frac{N_k^{-1/2}}{3!}\frac{\partial^3[A_k(\boldsymbol{y})P(\boldsymbol{y};t)]}{\partial w_k^3} + \cdots \Big\}. \tag{D.16}$$

In Eq. (D.15), $\boldsymbol{\mu}(t)$ solves the macroscopic differential equations:

$$\frac{d\mu_k(t)}{dt} = [1-\mu_k(t)]\big[\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{\mu}(t)))\big] - \mu_k(t)\big[\lambda_k^- + \gamma_k(\rho_k(\boldsymbol{\mu}(t)))\big], \ t > 0, \ k \in \mathcal{K}, \tag{D.17}$$

initialized by $\boldsymbol{\mu}(0) = \mathbf{0}$, whereas, for each t, $W_k(t)$, $k \in \mathcal{K}$, are zero-mean correlated Gaussian random variables. We can now use a Taylor expansion of the drift and diffusion functions to obtain:

$$\begin{aligned} A_k(\boldsymbol{y}) &= A_k\Big(\mu_1 + \frac{1}{\sqrt{N_1}}\, w_1, \ldots, \mu_K + \frac{1}{\sqrt{N_K}}\, w_K\Big) \\ &= A_k(\boldsymbol{\mu}) + \sum_{k'\in\mathcal{K}} N_{k'}^{-1/2} w_{k'} A_{kk'}(\boldsymbol{\mu}) + \sum_{k'\in\mathcal{K}}\sum_{k''\in\mathcal{K}} N_{k'}^{-1/2} N_{k''}^{-1/2} w_{k'} w_{k''} A_{kk'k''}(\boldsymbol{\mu}) + \cdots, \end{aligned} \tag{D.18}$$

for $k \in \mathcal{K}$, where $A_{kk'}(\boldsymbol{y}) := \partial A_k(\boldsymbol{y})/\partial y_{k'}$ and $A_{kk'k''}(\boldsymbol{y}) := \partial^2 A_k(\boldsymbol{y})/\partial y_{k'}\partial y_{k''}$, and likewise for the diffusion functions. In this case, Eq. (D.16) becomes

$$\begin{aligned} \frac{\partial P(\boldsymbol{y};t)}{\partial t} = &-\sum_{k\in\mathcal{K}} N_k^{1/2}\frac{\partial[A_k(\boldsymbol{\mu})P(\boldsymbol{y};t)]}{\partial w_k} - \sum_{k\in\mathcal{K}} N_k^{1/2}\sum_{k'\in\mathcal{K}} N_{k'}^{-1/2}\frac{\partial[w_{k'}A_{kk'}(\boldsymbol{\mu})P(\boldsymbol{y};t)]}{\partial w_k} \\ &+\frac{1}{2}\sum_{k\in\mathcal{K}}\frac{\partial^2[D_k(\boldsymbol{\mu})P(\boldsymbol{y};t)]}{\partial w_k^2} + \mathcal{O}(N_k^{-1/2}). \end{aligned} \tag{D.19}$$

APPENDIX D

We can now replace the probability distribution $P(\boldsymbol{y};t)$ of the fractional activity process with the probability distribution $P'(\boldsymbol{w};t)$ of the noise process, in which case we obtain

$$\begin{aligned}\frac{\partial P'(\boldsymbol{w};t)}{\partial t} &= \sum_{k\in\mathcal{K}} N_k^{1/2}\left[\frac{d\mu_k(t)}{dt} - A_k(\boldsymbol{\mu}(t))\right]\frac{\partial P'(\boldsymbol{w};t)}{\partial w_k} \\ &\quad - \sum_{k\in\mathcal{K}} N_k^{1/2}\sum_{k'\in\mathcal{K}} N_{k'}^{-1/2} A_{kk'}(\boldsymbol{\mu}(t))\,\frac{\partial[w_{k'}P'(\boldsymbol{w};t)]}{\partial w_k} \\ &\quad + \frac{1}{2}\sum_{k\in\mathcal{K}} D_k(\boldsymbol{\mu}(t))\,\frac{\partial^2 P'(\boldsymbol{w};t)}{\partial w_k^2} + \mathcal{O}(N_k^{-1/2}),\end{aligned} \tag{D.20}$$

by virtue of the fact that

$$\begin{aligned}\frac{\partial P(\boldsymbol{y};t)}{\partial t} &= \frac{\partial P'(\boldsymbol{w};t)}{\partial t} + \sum_{k\in\mathcal{K}}\frac{dw_k(t)}{dt}\frac{\partial P'(\boldsymbol{w};t)}{\partial w_k} \\ &= \frac{\partial P'(\boldsymbol{w};t)}{\partial t} - \sum_{k\in\mathcal{K}} N_k^{1/2}\frac{d\mu_k(t)}{dt}\frac{\partial P'(\boldsymbol{w};t)}{\partial w_k},\end{aligned} \tag{D.21}$$

where the second equality is a consequence of the fact that $dY_k(t)/dt = 0$ (except at time points on a set of measure zero at which the derivative is infinite). This implies that $dw_k(t)/dt = -N_k^{1/2}d\mu_k(t)/dt$, by virtue of Eq. (D.15).

Note now that $\boldsymbol{\mu}(t)$ satisfies the macroscopic equations $d\mu_k(t)/dt = A_k(\boldsymbol{\mu}(t))$. As a consequence, and by virtue of Eq. (D.20), the noise probability distribution $P'(\boldsymbol{w};t)$ is approximately governed by the *linear* Fokker-Planck equation

$$\frac{\partial P'(\boldsymbol{w};t)}{\partial t} = -\sum_{k\in\mathcal{K}}\sum_{k'\in\mathcal{K}}\sqrt{\frac{\zeta_k}{\zeta_{k'}}}\,A_{kk'}(\boldsymbol{\mu}(t))\,\frac{\partial[w_{k'}P'(\boldsymbol{w};t)]}{\partial w_k} + \frac{1}{2}\sum_{k\in\mathcal{K}} D_k(\boldsymbol{\mu}(t))\,\frac{\partial^2 P'(\boldsymbol{w};t)}{\partial w_k^2}, \tag{D.22}$$

where $\zeta_k = N_k/N$ and we ignore the terms of $\mathcal{O}(N_k^{-1/2})$. Note that this equation does not depend on N, since we have used the relation $N_k/N_{k'} = \zeta_k/\zeta_{k'}$, which is true at any point *en route* to the thermodynamic limit. The solution to this equation is a multivariate Gaussian density with zero mean and correlation matrix $\mathbb{R}(t)$ with elements $r_{kk'}(t) = \mathrm{E}[W_k(t)W_{k'}(t)]$ that satisfy the following system of Lyapunov equations (5.2.47), initialized with $r_{kk'}(0) = 0$, for every $k, k' \in \mathcal{K}$.

We note here that there are two approximation steps involved with the LNA method. The first step is the *ansatz* given by Eq. (D.15), whereas the second step is ignoring all terms of

$\mathcal{O}(N_k^{-1/2})$ in Eq. (D.20). Theoretically speaking, justification of the second step is simple – we can take N to be large so that $1/\sqrt{N_k}$ is small enough that any term in the expansion multiplied by $1/\sqrt{N_k}$ is negligible. This requirement however is not useful in practice, since we cannot determine the appropriate value of N that satisfies the required condition. On the other hand, assuming that N is large enough for the previous approximation to be valid, Eq. (D.15) can be justified only for *monostable* systems [1].

Leakiness and Irreducibility

In the theory of Markov processes, the state $\boldsymbol{y}$ is said to be *accessible* from another state $\boldsymbol{y}'$ if there is a non-zero probability to transition (possibly through intermediate states) from $\boldsymbol{y}'$ to $\boldsymbol{y}$. We denote this by $\boldsymbol{y}' \to \boldsymbol{y}$. The states $\boldsymbol{y}$ and $\boldsymbol{y}'$ are said to be *communicating* whenever $\boldsymbol{y} \to \boldsymbol{y}'$ and $\boldsymbol{y}' \to \boldsymbol{y}$. In this case, we write $\boldsymbol{y} \rightleftarrows \boldsymbol{y}'$. If all states $\boldsymbol{y} \in \mathcal{Y}$ are communicating, then there is a non-zero probability to transition from any state to any other state, and the Markov process is said to be *irreducible*. An irreducible Markov process has a unique, strictly positive, and asymptotically stable stationary probability distribution that is independent of the initial condition.

It is often difficult to prove that a Markov processes is irreducible. The following results allow us to determine when the fractional activity process $\boldsymbol{Y}(t)$ in a LMN is irreducible. It turns out that leakiness is intimately related to irreducibility.

Proposition 1. *If $\lambda_k^+ > 0$, for every $k \in \mathcal{K}$, then $\mathbf{0} \to \boldsymbol{y}' \to \mathbf{1}$ for all $\boldsymbol{y}' \in \mathcal{Y}$, where $\mathbf{0}$ is the state of zero fractional inactivity and $\mathbf{1}$ is the state of maximum fractional activity.*

Proof. For every $\boldsymbol{y} \in \mathcal{Y}$ such that $y_k \leq 1 - 1/N_k$, for some $k \in \mathcal{K}$, we have $\pi_k^+(\boldsymbol{y}) = N_k(1 - y_k)[\lambda_k^+ + \phi_k(\rho_k(\boldsymbol{y}))] \geq \lambda_k^+ + \phi_k(\rho_k(\boldsymbol{y})) > 0$, where the second inequality is due to the fact that $\lambda_k^+ > 0$ and $\phi_k(\rho_k(\boldsymbol{y})) \geq 0$. Therefore, $\boldsymbol{y} \to \boldsymbol{y} + \widetilde{\boldsymbol{e}}_k$, where $\widetilde{\boldsymbol{e}}_k$ is the k-th column of the $K \times K$ identity matrix multiplied by N_k^{-1}. Thus, for any state $\boldsymbol{y} \in \mathcal{Y}$ such that $\boldsymbol{y} + \widetilde{\boldsymbol{e}}_k \in \mathcal{Y}$, we have that $\boldsymbol{y} \to \boldsymbol{y} + \widetilde{\boldsymbol{e}}_k$. Let $\boldsymbol{y}' = (n_1/N_1, n_2/N_2, \ldots, n_K/N_K)^T$, for some $n_1, n_2, \ldots, n_K$. A non-zero probability path from $\mathbf{0}$ to $\boldsymbol{y}'$

is then given by the following sequence: n_1 transitions taking $\boldsymbol{y} \to \boldsymbol{y} + \widetilde{\boldsymbol{e}}_1$, followed by n_2 transitions taking $\boldsymbol{y} \to \boldsymbol{y} + \widetilde{\boldsymbol{e}}_2$,..., followed by n_K transitions taking $\boldsymbol{y} \to \boldsymbol{y} + \widetilde{\boldsymbol{e}}_K$. Likewise, a non-zero probability path from $\boldsymbol{y}'$ to $\mathbf{1}$ is given by the following sequence: $N_1 - n_1$ transitions taking $\boldsymbol{y} \to \boldsymbol{y} + \widetilde{\boldsymbol{e}}_1$, followed by $N_2 - n_2$ transitions taking $\boldsymbol{y} \to \boldsymbol{y} + \widetilde{\boldsymbol{e}}_2$,..., followed by $N_K - n_K$ transitions taking $\boldsymbol{y} \to \boldsymbol{y} + \widetilde{\boldsymbol{e}}_K$. Hence, $\mathbf{0} \to \boldsymbol{y}' \to \mathbf{1}$, for all $\boldsymbol{y}' \in \mathcal{Y}$. □

Corollary 1. *If $\lambda_k^+ > 0$, for every $k \in \mathcal{K}$, and $\mathbf{1} \to \mathbf{0}$, then $\boldsymbol{Y}(t)$ is irreducible.*

Proof. From Proposition 1, and for any $\boldsymbol{y}', \boldsymbol{y}'' \in \mathcal{Y}$, we have that $\boldsymbol{y}' \to \mathbf{1}$ and $\mathbf{0} \to \boldsymbol{y}''$. Since, $\mathbf{1} \to \mathbf{0}$, this implies $\boldsymbol{y}' \to \mathbf{1} \to \mathbf{0} \to \boldsymbol{y}''$ and thus $\boldsymbol{Y}(t)$ is irreducible. □

Proposition 2. *If $\lambda_k^- > 0$, for every $k \in \mathcal{K}$, then $\mathbf{1} \to \boldsymbol{y}' \to \mathbf{0}$ for all $\boldsymbol{y}' \in \mathcal{Y}$.*

Proof. For every $\boldsymbol{y} \in \mathcal{Y}$ such that $y_k \geq 1/N_k$, for some $k \in \mathcal{K}$, we have $\pi_k^-(\boldsymbol{y}) = N_k y_k[\lambda_k^- + \gamma_k(\rho_k(\boldsymbol{y}))] \geq \lambda_k^- + \gamma_k(\rho_k(\boldsymbol{y})) > 0$, where the second inequality is due to the fact that $\lambda_k^- > 0$ and $\gamma_k(\rho_k(\boldsymbol{y})) \geq 0$. Therefore, $\boldsymbol{y} \to \boldsymbol{y} - \widetilde{\boldsymbol{e}}_k$. Thus, for any state $\boldsymbol{y} \in \mathcal{Y}$ such that $\boldsymbol{y} - \widetilde{\boldsymbol{e}}_k \in \mathcal{Y}$, we have that $\boldsymbol{y} \to \boldsymbol{y} - \widetilde{\boldsymbol{e}}_k$. Let $\boldsymbol{y}' = (n_1/N_1, n_2/N_2, \ldots, n_K/N_K)^T$, for some $n_1, n_2, \ldots, n_K$. A non-zero probability path from $\mathbf{1}$ to $\boldsymbol{y}'$ is then given by the following sequence: $N_1 - n_1$ transitions taking $\boldsymbol{y} \to \boldsymbol{y} - \widetilde{\boldsymbol{e}}_1$, followed by $N_2 - n_2$ transitions taking $\boldsymbol{y} \to \boldsymbol{y} - \widetilde{\boldsymbol{e}}_2$,..., followed by $N_K - n_K$ transitions taking $\boldsymbol{y} \to \boldsymbol{y} - \widetilde{\boldsymbol{e}}_K$. On the other hand, a non-zero probability path from $\boldsymbol{y}'$ to $\mathbf{0}$ is given by the following sequence: n_1 transitions taking $\boldsymbol{y} \to \boldsymbol{y} - \widetilde{\boldsymbol{e}}_1$, followed by n_2 transitions taking $\boldsymbol{y} \to \boldsymbol{y} - \widetilde{\boldsymbol{e}}_2$,..., followed by n_K transitions taking $\boldsymbol{y} \to \boldsymbol{y} - \widetilde{\boldsymbol{e}}_K$. Hence, $\mathbf{1} \to \boldsymbol{y}' \to \mathbf{0}$, for all $\boldsymbol{y}' \in \mathcal{Y}$. □

Corollary 2. *If $\lambda_k^- > 0$, for every $k \in \mathcal{K}$, and $\mathbf{0} \to \mathbf{1}$, then $\boldsymbol{Y}(t)$ is irreducible.*

Proof. From Proposition 2, and for any $\boldsymbol{y}', \boldsymbol{y}'' \in \mathcal{Y}$, we have that $\boldsymbol{y}' \to \mathbf{0}$ and $\mathbf{1} \to \boldsymbol{y}''$. Since, $\mathbf{0} \to \mathbf{1}$, this implies $\boldsymbol{y}' \to \mathbf{0} \to \mathbf{1} \to \boldsymbol{y}''$ and thus $\boldsymbol{Y}(t)$ is irreducible. □

By combining the previous results, we obtain the following theorem:

Theorem 1. *If $\lambda_k^-, \lambda_k^+ > 0$, for every $k \in \mathcal{K}$, then $\boldsymbol{Y}(t)$ is irreducible.*

APPENDIX D

Defining Avalanches

A common way to define avalanches has originated from work on neural networks, since the emergence of avalanches is a fundamental property of such networks [4, 168]. This definition is based on partitioning time into bins $\mathcal{T}_i = [i\Delta t, (i+1)\Delta t)$, $i = 0, 1, \ldots$, of equal duration $\Delta t > 0$ and associating to each bin $\mathcal{T}_i$ a frame $\mathcal{F}_i$, defined as the portion $\{\boldsymbol{X}(t), t \in \mathcal{T}_i\}$ of the activity process during $\mathcal{T}_i$. The frame $\mathcal{F}_i$ is said to be blank if the activity process $\boldsymbol{X}(t)$ is zero within $\mathcal{T}_i$; otherwise, the frame $\mathcal{F}_i$ is said to be active. Then, an avalanche is defined to be a sequence of consecutively active frames that is preceded and ended by a blank frame [168]. Note that, if

$$\mathcal{W}(t) := \frac{1}{N} \sum_{n \in \mathcal{N}} X_n(t) = \sum_{k \in \mathcal{K}} \zeta_k Y_k(t) \tag{D.23}$$

is the net fractional activity of the activity process $\boldsymbol{X}(t)$ at time t, where $\zeta_k = N_k/N$, then $\mathcal{F}_i$ is blank if $\max_{t \in \mathcal{T}_i}\{\mathcal{W}(t)\} \leq \epsilon$, whereas $\mathcal{F}_i$ is active if $\max_{t \in \mathcal{T}_i}\{\mathcal{W}(t)\} > \epsilon$. Here, $\epsilon \geq 0$ is a small threshold that dictates the minimum percentage of nodes that can be active in the network in order for the network to be deemed active.

Unfortunately, the previous definition depends on the choice of Δt. Moreover, the definition is sensitive to arbitrary shifts of the time axis. For this reason, we provide here an alternative definition for an avalanche that is not influenced by the previous factors. In particular, we say that an avalanche occurs within a time window $[t, t+\tau)$, whenever the following three conditions are satisfied: (i) There exist some small $dt > 0$ such that $\mathcal{W}(t') \leq \epsilon$, for all $t' \in [t - dt, t)$; i.e., the network is inactive immediately before time t. (ii) $\mathcal{W}(t') > \epsilon$, for all $t' \in [t, t+\tau)$; i.e., the network is active during the time interval $[t, t+\tau)$. (iii) $\mathcal{W}(t+\tau) \leq \epsilon$; i.e., the network becomes inactive at time $t + \tau$.

It is common to characterize an avalanche by two parameters: its duration and size. The duration d_a of an avalanche has been defined to be the number of consecutively active frames multiplied by Δt. On the other hand, its size s_a is simply the number of times within the duration that an element of the activity process $\boldsymbol{X}(t)$ becomes active (switches from 0 to 1). Using our

definition of an avalanche, it is not difficult to see that $d_a = \tau$, whereas s_a is the number of times during $[t, t+\tau)$ that an element of the activity process $\boldsymbol{X}(t)$ becomes active (switches from 0 to 1). Note that a given element of the activity process may transition from 0 to 1 multiple times throughout the duration. As a consequence, the size of an avalanche is not limited by N. To account for the effect of population size, it is common to consider the fractional avalanche size s_a/N [4].

Bibliography

[1] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. Amsterdam: Elsevier, 2007.

[2] M. Cook, D. Soloveichik, E. Winfree, and J. Bruck, "Programmability of chemical reaction networks," in *Algorithmic Bioprocesses*, ser. Natural Computing Series, A. Condon, D. Harel, J. N. Kok, A. Salomaa, and E. Winfree, Eds. Berlin: Springer-Verlag, 2009, pp. 543–584.

[3] M. O. Magnasco, "Chemical kinetics is turing universal," *Phys. Rev. Lett.*, vol. 78, no. 6, pp. 1190–1193, 1997.

[4] A. Klaus, S. Yu, and D. Plenz, "Statistical analyses support power law distributions found in neuronal avalanches," *PLoS One*, vol. 6, no. 5, p. e19779, 2011.

[5] J. Goutsias and G. Jenkinson, "Markovian dynamics on complex reaction networks," *Phys. Rep.*, vol. 529, no. 2, pp. 199–264, 2013.

[6] S. Klamt, U.-U. Haus, and F. Theis, "Hypergraphs and cellular networks," *PLoS Comput. Biol.*, vol. 5:e1000385, no. 5, 2009.

[7] M. E. J. Newman, *Networks: An Introduction*. New York: Oxford University Press, 2010.

[8] J. Goutsias, "Classical versus stochastic kinetics modeling of biochemical reaction systems," *Biophys. J.*, vol. 92, pp. 2350–2365, 2007.

[9] S. M. Ross, *Stochastic Processes*, 2nd ed. New York: Wiley, 1996.

[10] D. T. Gillespie, "The chemical Langevin equation," *J. Chem. Phys.*, vol. 113, no. 1, pp. 297–306, 2000.

[11] E. L. Haseltine and J. B. Rawlings, "Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics," *J. Chem. Phys.*, vol. 117, no. 15, pp. 6959–6969, 2002.

[12] J. Goutsias, "Quasiequilibrim approximation of fast reaction kinetics in stochastic biochemical systems," *J. Chem. Phys.*, vol. 122:184102, 2005.

[13] ——, "A hidden Markov model for transcriptional regulation in single cells," *IEEE/ACM Trans. Comput. Biol. Boinf.*, vol. 3, no. 1, pp. 57–71, 2006.

[14] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A*, vol. 188, pp. 404–425, 1992.

[15] K. A. Dill and S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience*, 2nd ed. New York: Galand Science, 2011.

[16] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Rev.*, vol. 42, no. 4, pp. 599–653, 2000.

[17] W. Y. Chen and S. Bokka, "Stochastic modeling of nonlinear epidemiology," *J. Theor. Biol.*, vol. 234, pp. 455–470, 2005.

[18] Y. Ben-Zion, Y. Cohen, and N. M. Shnerb, "Modeling epidemics dynamics on heterogeneous networks," *J. Theor. Biol.*, vol. 264, pp. 197–204, 2010.

[19] M. Benayoun, J. D. Cowan, W. van Drongelen, and E. Wallace, "Avalanches in a stochastic model of spiking neurons," *PLoS Comput. Biol.*, vol. 6:e1000846, no. 7, 2010.

[20] D. A. McQuarrie, "Kinetics of small systems. I," *J. Chem. Phys.*, vol. 38, no. 2, pp. 433–436, 1963.

[21] D. A. McQuarrie, C. J. Jachimowski, and M. E. Russell, "Kinetics of small systems. II," *J. Chem. Phys.*, vol. 40, no. 10, pp. 2914–2921, 1964.

[22] I. G. Darvey and P. J. Staff, "Stochastic approach to first-order chemical reaction kinetics," *J. Chem. Phys.*, vol. 44, no. 3, pp. 990–997, 1966.

[23] D. Leonard and L. E. Reichl, "Stochastic analysis of a driven chemical reaction," *J. Chem. Phys.*, vol. 92, no. 10, pp. 6004–6010, 1990.

[24] I. J. Laurenzi, "An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics," *J. Chem. Phys.*, vol. 113, no. 8, pp. 3315–3322, 2000.

[25] C. Gadgil, C. H. Lee, and H. G. Othmer, "A stochastic analysis of first-order reaction networks," *B. Math. Biol.*, vol. 67, pp. 901–946, 2005.

[26] X. Zhang, K. De Cock, M. F. Bugallo, and P. M. Djurić, "A general method for the computation of probabilities in systems of first order chemical reactions," *J. Chem. Phys.*, vol. 122:104101, 2005.

[27] W. J. Heuett and H. Qian, "Grand canonical Markov model: A stochastic theory for open nonequilibrium biochemical networks," *J. Chem. Phys.*, vol. 124:044110, 2006.

[28] T. Jahnke and W. Huisinga, "Solving the chemical master equation for monomolecular reaction systems analytically," *J. Math. Biol.*, vol. 54, pp. 1–26, 2007.

[29] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. Berlin: Springer, 2010.

[30] B. Munsky and M. Khammash, "The finite state projection algorithm for the solution of the chemical master equation," *J. Chem. Phys.*, vol. 124:044104, 2006.

[31] C. Moler and C. van Loan, "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later," *SIAM Rev.*, vol. 45, no. 1, pp. 3–49, 2003.

[32] R. B. Sidje, "Expokit: A software package for computing matrix exponentials," *ACM T. Math. Software*, vol. 24, no. 1, pp. 130–156, 1998.

[33] R. B. Sidje and W. J. Stewart, "A numerical study of large sparse matrix exponentials arising in Markov chains," *Comput. Stat. Data An.*, vol. 29, pp. 345–368, 1999.

[34] S. Peleš, B. Munsky, and M. Khammash, "Reduction and solution of the chemical master equation using time scale separation and finite state projection," *J. Chem. Phys.*, vol. 125:204104, 2006.

[35] M. Hegland, C. Burden, L. Santoso, S. MacNamara, and H. Booth, "A solver for the stochastic master equation applied to gene regulatory networks," *J. Comput. Appl. Math.*, vol. 205, pp. 708–724, 2007.

[36] B. Munsky and M. Khammash, "A multiple time interval finite state projection algorithm for the solution to the chemical master equation," *J. Comput. Phys.*, vol. 226, pp. 818–835, 2007.

[37] P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow, "Adaptive discrete Galerkin methods applied to the chemical master equation," *SIAM J. Sci. Comput.*, vol. 30, no. 6, pp. 2990–3011, 2008.

[38] M. Hegland, A. Hellander, and P. Lötstedt, "Sparse grids and hybrid methods for the chemical master equation," *BIT*, vol. 48, pp. 265–283, 2008.

[39] T. Jahnke and W. Huisinga, "A dynamical low-rank approach to the chemical master equation," *B. Math. Biol.*, vol. 70, pp. 2283–2302, 2008.

[40] S. MacNamara, K. Burrage, and R. B. Sidje, "Multiscale modeling of chemical kinetics via the master equation," *Multiscale Model. Simul.*, vol. 6, no. 4, pp. 1146–1168, 2008.

[41] V. Wolf, R. Goel, M. Mateescu, and T. A. Henzinger, "Solving the chemical master equation using sliding windows," *BMC Syst. Biol.*, vol. 4:42, 2010.

[42] J. Zhang, L. T. Watson, and Y. Cao, "A modified uniformization method for the solution of the chemical master equation," *Comp. Math. Appl.*, vol. 59, pp. 573–584, 2010.

[43] T. Jahnke, "An adaptive wavelet method for the chemical master equation," *SIAM J. Sci. Comput.*, vol. 31, no. 6, pp. 4373–4394, 2010.

[44] T. Jahnke and T. Udrescu, "Solving chemical master equations by adaptive wavelet compression," *J. Comput. Phys.*, vol. 229, pp. 5724–5741, 2010.

[45] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comput. Phys.*, vol. 22, pp. 403–434, 1976.

[46] ——, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, no. 25, pp. 2340–2361, 1977.

[47] ——, "Simulation methods in systems biology," in *Formal Methods for Computational Systems Biology*, ser. Lecture Notes in Computer Science, M. Bernardo, P. Degano, and G. Zavattaro, Eds. Berlin: Springer-Verlag, 2008, vol. 5016, pp. 125–167.

[48] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *J. Phys. Chem. A*, vol. 104, pp. 1876–1889, 2000.

[49] Y. Cao, H. Li, and L. Petzold, "Efficient formulation of the stochastic simulation algorithm for chemically reacting systems," *J. Chem. Phys.*, vol. 121, no. 9, pp. 4059–4067, 2004.

[50] J. M. McCollum, G. D. Peterson, C. D. Cox, M. L. Simpson, and N. F. Samatova, "The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior," *Comput. Biol. Chem.*, vol. 30, no. 1, pp. 39–49, 2006.

[51] X. Cai and Z. Xu, "k-leap method for accelerating stochastic simulation of coupled chemical reactions," *J. Chem. Phys.*, vol. 126:074102, 2007.

[52] A. Lipshtat, ""All possible steps" approach to the accelerated use of Gillespie's algorithm," *J. Chem. Phys.*, vol. 126:184103, 2007.

[53] A. Hellander, "Efficient computation of transient solutions of the chemical master equation based on uniformization and quasi-Monte Carlo," *J. Chem. Phys.*, vol. 128:154109, 2008.

[54] A. Slepoy, A. P. Thompson, and S. J. Plimpton, "A constant-time kinetic Monte Carlo algorithm for simulation of large biochemical reaction networks," *J. Chem. Phys.*, vol. 128:205101, 2008.

[55] X. Cai and J. Wen, "Efficient exact and k-skip methods for stochastic simulation of coupled chemical reactions," *J. Chem. Phys.*, vol. 131:064108, 2009.

[56] E. Mjolsness, D. Orendorff, P. Chatelain, and P. Koumoutsakos, "An exact accelerated stochastic simulation algorithm," *J. Chem. Phys.*, vol. 130:144110, 2009.

[57] R. Ramaswamy, N. González-Segredo, and I. F. Sbalzarini, "A new class of highly efficient exact stochastic simulation algorithms for chemical networks," *J. Chem. Phys.*, vol. 130:244104, 2009.

[58] S. Wu, J. Fu, Y. Cao, and L. Petzold, "Michaelis-Menten speeds up tau-leaping under a wide range of conditions," *J. Chem. Phys.*, vol. 134:134112, 2011.

[59] T. G. Kurtz, "Representations of Markov processes as multiparameter time changes," *Ann. Probab.*, vol. 8, no. 4, pp. 682–715, 1980.

[60] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *J. Chem. Phys.*, vol. 115, no. 4, pp. 1716–1733, 2001.

[61] D. T. Gillespie and L. R. Petzold, "Improved leap-size selection for accelerated stochastic simulation," *J. Chem. Phys.*, vol. 119, no. 16, pp. 8229–8234, 2003.

[62] Y. Cao, D. T. Gillespie, and L. R. Petzold, "Efficient step size selection for the tau-leaping simulation method," *J. Chem. Phys.*, vol. 124:044109, 2006.

[63] T. Tian and K. Burrage, "Binomial leap methods for simulating stochastic chemical kinetics," *J. Chem. Phys.*, vol. 121, no. 21, pp. 10 356–10 364, 2004.

[64] A. Chatterjee, K. Mayawala, J. S. Edwards, and D. G. Vlachos, "Time accelerated Monte Carlo simulations of biological networks using the binomial τ-leap method," *Bioinformatics*, vol. 21, no. 9, pp. 2136–2137, 2005.

[65] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis, "Binomial distribution based τ-leap accelerated stochastic simulation," *J. Chem. Phys.*, vol. 122:024112, 2005.

[66] X. Peng, W. Zhou, and Y. Wang, "Efficient binomial leap method for simulating chemical kinetics," *J. Chem. Phys.*, vol. 126:224109, 2007.

[67] M. F. Pettigrew and H. Resat, "Multinomial tau-leaping method for stochastic kinetic simulations," *J. Chem. Phys.*, vol. 126:084101, 2007.

[68] Z. Xu and X. Cai, "Unbiased τ-leap methods for stochastic simulation of chemically reacting systems," *J. Chem. Phys.*, vol. 128:154112, 2008.

[69] Y. Xu and Y. Lan, "The n-leap method for stochastic simulation of coupled chemical reactions," *J. Chem. Phys.*, vol. 137, p. 204103, 2012.

[70] D. Orendorff and E. Mjolsness, "A hierarchical exact accelerated stochastic simulation algorithm," *J. Chem. Phys.*, vol. 137, p. 214104, 2012.

[71] Y. Cao, D. T. Gillespie, and L. R. Petzold, "Avoiding negative populations in explicit poisson tau-leaping," *J. Chem. Phys.*, vol. 123:054104, 2005.

[72] D. F. Anderson, "Incorporating postleap checks in tau-leaping," *J. Chem. Phys.*, vol. 128:054103, 2008.

[73] B. Mélykúti, "Theoretical advances in the modelling and interrogation of biochemical reaction systems: Alternative formulations of the chemical langevin equation and optimal experiment design for model discrimination," (http://ora.ox.ac.uk/objects/uuid:d368c04c-b611-41b2-8866-cde16b283b0d), Keble College, University of Oxford, Oxford, England, 2010.

[74] T. G. Kurtz, "Limit theorems for sequences of jump Markov processes approximating ordinary differential processes," *J. Appl. Prob.*, vol. 8, pp. 344–356, 1971.

[75] ——, "The relationship between stochastic and deterministic models for chemical reactions," *J. Chem. Phys.*, vol. 57, no. 7, pp. 2976–2978, 1972.

[76] N. G. van Kampen, "A power series expansion of the master equation," *Can. J. Phys.*, vol. 39, pp. 551–567, 1961.

[77] ——, "The expansion of the master equation," in *Advance in Chemical Physics*, I. Prigogine and S. A. Rice, Eds. New York: John Wiley & Sons, 1976, vol. 34, pp. 245–309.

[78] J. Elf and M. Ehrenberg, "Fast evaluation of fluctuations in biochemical networks with the linear noise approximation," *Genome Res.*, vol. 13, pp. 2475–2484, 2003.

[79] J. Elf, J. Paulsson, O. G. Berg, and M. Ehrenberg, "Near-critical phenomena in intracellular metabolite pools," *Biophys. J.*, vol. 84, pp. 154–170, 2003.

[80] F. Hayot and C. Jayaprakash, "The linear noise approximation for molecular fluctuations within cells," *Phys. Biol.*, vol. 1, pp. 205–210, 2004.

[81] R. Tomioka, H. Kimura, T. J. Kobayashi, and K. Aihara, "Multivariate analysis of noise in genetic regulatory networks," *J. Theor. Biol.*, vol. 229, pp. 501–521, 2004.

[82] Y. Tao, Y. Jia, and T. G. Dewey, "Stochastic fluctuations in gene expression far from equilibrium: Ω expansion and linear noise approximation," *J. Chem. Phys.*, vol. 122:124108, 2005.

[83] M. Scott, B. Ingalls, and M. Kaern, "Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks," *Chaos*, vol. 16:026107, 2006.

[84] A. J. McKane, J. D. Nagy, T. J. Newman, and M. O. Stefanini, "Amplified biochemical oscillations in cellular systems," *J. Stat. Phys.*, vol. 128, pp. 165–191, 2007.

[85] T. Dauxois, F. Di Patti, D. Fanelli, and A. J. McKane, "Enhanced stochastic oscillations in autocatalytic reactions," *Phys. Rev. E*, vol. 79, p. 036112, 2009.

[86] A. J. McKane and T. J. Newman, "Predator-prey cycles from resonant amplification of demographic stochasticity," *Phys. Rev. Lett.*, vol. 94, p. 218102, 2005.

[87] S. Datta, G. W. Delius, and R. Law, "A jump-growth model for predator-pray dynamics: Derivation and application to marine ecosystems," *B. Math. Biol.*, vol. 72, pp. 1361–1382, 2010.

[88] M. S. de la Lama, I. G. Szendro, J. R. Iglesias, and H. S. Wio, "Van Kampen's expansion approach in an opinion formation model," *Eur. Phys. J. B*, vol. 51, pp. 435–442, 2006.

[89] P. C. Bressloff, "Stochastic neural field theory and the system-size expansion," *SIAM J. Appl. Math.*, vol. 70, no. 5, pp. 1488–1521, 2009.

[90] M. M. Mansour, C. van Den Broeck, G. Nicolis, and J. W. Turner, "Asymptotic properties of Markovian master equations," *Ann. Phys.*, vol. 131, pp. 283–313, 1981.

[91] J. Schnakenberg, "Network theory of microscopic and macroscopic behavior of master equation systems," *Rev. Mod. Phys.*, vol. 48, no. 4, pp. 571–585, 1976.

[92] M. J. Keeling and J. V. Ross, "On methods for studying stochastic disease dynamics," *J. R. Soc. Interface*, vol. 5, pp. 171–181, 2008.

[93] M. Vellela and H. Qian, "Stochastic dynamics and non-equilibrium thermodynamics of a

bistable chemical system: The Schlögl model revisited," *J. R. Soc. Interface*, vol. 6, pp. 925–940, 2009.

[94] H. Gang, "Lyapunov function and stationary probability distributions," *Z. Phys. B – Condensed Matter*, vol. 65, pp. 103–106, 1986.

[95] W. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control: A Lyapunov-based Approach*. Princeton: Princeton University Press, 2008.

[96] P. Ao, C. Kwon, and H. Qian, "On the existence of potential landscape in the evolution of complex systems," *Complexity*, vol. 12, no. 4, pp. 19–27, 2007.

[97] J. Wang, L. Xu, and E. Wang, "Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations," *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 34, pp. 12 271–12 276, 2008.

[98] D. Zhou and H. Qian, "Fixation, transient landscape, and diffusion dilemma in stochastic evolutionary game dynamics," *Phys. Rev. E*, vol. 84:031907, 2011.

[99] J. Wang, K. Zhang, L. Xu, and E. Wang, "Quantifying the Waddington landscape and biological paths for development and differentiation," *Proc. Natl. Acad. Sci. USA*, vol. 108, no. 20, pp. 8257–8262, 2011.

[100] J. Keizer, *Statistical Thermodynamics of Nonequilibrium Processes*. New York: Springer-Verlag, 1987.

[101] M. Vellela and H. Qian, "A quasistationary analysis of a stochastic chemical reaction: Keizer's paradox," *B. Math. Biol.*, vol. 69, pp. 1727–1746, 2007.

[102] H. Qian, "Cellular biology in terms of stochastic nonlinear biochemical dynamics: Emergent properties, isogenetic variations and chemical system inheritability," *J. Stat. Phys.*, vol. 141, pp. 990–1013, 2010.

[103] M. N. Artyomov, J. Das, M. Kardar, and A. K. Chakraborty, "Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities," *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 48, pp. 18 958–18 963, 2007.

[104] M. N. Artyomov, M. Mathur, M. S. Samoilov, and A. K. Chakraborty, "Stochastic bimodalities in deterministically monostable reversible chemical networks due to network topology reduction," *J. Chem. Phys.*, vol. 131:195103, 2009.

[105] H. Qian, P.-Z. Shi, and J. Xing, "Stochastic bifurcation, slow fluctuations, and bistability as an origin of biochemical complexity," *Phys. Chem. Chem. Phys.*, vol. 11, pp. 4861–4870, 2009.

[106] L. M. Bishop and H. Qian, "Stochastic bistability and bifurcation in a mesoscopic signaling system with autocatalytic kinase," *Biophys. J.*, vol. 98, pp. 1–11, 2010.

[107] H. Qian, "Nonlinear stochastic dynamics of mesoscopic homogeneous biochemical reaction systems – An analytical theory," *Nonlinearity*, vol. 24, pp. R19–R49, 2011.

[108] D. J. Aldous, "Markov chains with almost exponential hitting times," *Stoch. Proc. Appl.*, vol. 13, pp. 305–310, 1982.

[109] I. Prigogine, "Time, structure, and fluctuations," *Science*, vol. 201, no. 4358, pp. 777–785, 1978.

[110] Y. Oono and M. Paniconi, "Steady state thermodynamics," *Prog. Theor. Phys. Supp.*, vol. 130, pp. 29–44, 1998.

[111] D. Andrieux and P. Gaspard, "Fluctuation theorem and Onsager reciprocity relations," *J. Chem. Phys.*, vol. 121, no. 13, pp. 6167–6174, 2004.

[112] P. Gaspard, "Fluctuation theorem for nonequilibrium reactions," *J. Chem. Phys.*, vol. 120, no. 19, pp. 8898–8905, 2004.

[113] D. Andrieux and P. Gaspard, "Fluctuation theorem for currents and Schnakenberg network theory," *J. Stat. Phys.*, vol. 127, no. 1, pp. 107–131, 2007.

[114] T. Schmiedl and U. Seifert, "Stochastic thermodynamics of chemical reaction networks," *J. Chem. Phys.*, vol. 126:044101, 2007.

[115] J. Ross, *Thermodynamics and Fluctuations far from Equilibrium.* Berlin: Springer-Verlag, 2008.

[116] H. Ge, "Extended forms of the second law for general time-dependent stochastic processes," *Phys. Rev. E*, vol. 80:021137, 2009.

[117] H. Qian, "Entropy demystified: The "thermo"-dynamics of stochastically fluctuating systems," in *Methods in Enzymology*, L. J. Michael and L. Brand, Eds. San Diego, CA: Elsevier, 2009, vol. 467, pp. 111–134.

[118] M. Esposito and C. van den Broeck, "Three detailed fluctuation theorems," *Phys. Rev. Lett.*, vol. 104:090601, 2010.

[119] ——, "Three faces of the second law. I. Master equation formulation," *Phys. Rev. E*, vol. 82:011143, 2010.

[120] H. Ge and H. Qian, "Physical origins of entropy production, free energy dissipation, and their mathematical representations," *Phys. Rev. E*, vol. 81:051133, 2010.

[121] J. Ross and A. F. Villaverde, "Thermodynamics and fluctuations far from equilibrium," *Entropy*, vol. 12, pp. 2199–2243, 2010.

[122] M. Santillán and H. Qian, "Irreversible thermodynamics in multiscale stochastic dynamical systems," *Phys. Rev. E*, vol. 83:041130, 2011.

[123] X.-J. Zhang, M. Qian, and H. Qian, "Stochastic theory of nonequilibrium steady states and it applications. Part I," *Phys. Rep.*, vol. 510, pp. 1–86, 2012.

[124] B. Han and J. Wang, "Least dissipation cost as a design principle for robustness and function of cellular networks," *Phys. Rev. E*, vol. 77:031922, 2008.

[125] H. Ge, H. Qian, and M. Qian, "Stochastic theory of nonequilibrium steady states. Part II: Applications in chemical biophysics," *Phys. Rep.*, vol. 510, pp. 87–118, 2012.

[126] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: John Wiley & Sons, 1991.

[127] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations.* New York: John Wiley & Sons, 1977.

[128] H. Haken, "Exact stationary solution of the master equation for systems far from thermal equilibrium in detailed balance," *Phys. Lett. A*, vol. 46, pp. 443–444, 1974.

[129] M. Youssef and C. Scoglio, "An individual-based approach to SIR epidemics in contact networks," *J. Theor. Biol.*, vol. 283, pp. 136–144, 2011.

[130] A. J. Black and A. J. McKane, "Stochasticity in staged models of epidemics: quantifying the dynamics of whooping cough," *J. R. Soc. Interface*, vol. 7, no. 49, pp. 1219–1227, 2010.

[131] M. J. Keeling and J. V. Ross, "Efficient methods for studying stochastic disease and population dynamics," *Theor. Popul. Biol.*, vol. 75, pp. 133–141, 2009.

[132] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. New York: Cambridge University Press, 2007.

[133] G. C. Fox, R. D. Williams, and P. C. Messina, *Parallel Computing Works!* San Francisco, CA: Morgan Kaufmann Publishers, Inc., 1994.

[134] I. Faragó, Á. Havasi, and Z. Zlatev, "Efficient implementation of stable Richardson extrapolation algorithms," *Comput. Math. Appl.*, vol. 60, no. 8, pp. 2309–2325, 2010.

[135] CDSC & CDU, "Influenza in a boarding school," *Brit. Med. J.*, vol. 1, no. 6112, p. 587, 1978.

[136] J. D. Murray, *Mathematical Biology. I. An Introduction*, 3rd ed. New York, NY: Springer-Verlag, 2001.

[137] E. L. Haseltine and J. B. Rawlings, "On the origins of approximations for stochastic chemical kinetics," *J. Chem. Phys.*, vol. 123:164115, 2005.

[138] J. Mayer, "Parallel algorithms for solving linear systems with sparse triangular matrices," *Computing*, vol. 86, pp. 291–312, 2009.

[139] K. N. Crank, "A method for approximating the probability functions of a Markov chain," *J. Appl. Prob.*, vol. 25, pp. 808–814, 1988.

[140] N. C. Severo, "A recursion theorem on solving differential-difference equations and applications to some stochastic processes," *J. Appl. Prob.*, vol. 6, pp. 673–681, 1969.

[141] R. J. Kryscio and N. C. Severo, "Computational and estimation procedures in multidimensional right-shift processes and some applications," *Adv. Appl. Prob.*, vol. 7, pp. 349–382, 1975.

[142] G. Lente, "Stochastic mapping of first order reaction networks: A systematic comparison of the stochastic and deterministic kinetic approaches," *J. Chem. Phys.*, vol. 137, pp. 1–8, 2012.

[143] Y. Cao and L. Petzold, "Accuracy limitations and the measurement of errors in the stochastic simulation of chemically reacting systems," *J. Comp. Phys.*, vol. 212, pp. 6–24, 2006.

[144] R. Gunawan, Y. Cao, L. Petzold, and F. J. Doyle III, "Sensitivity analysis of discrete stochastic systems," *Biophys. J.*, vol. 88, pp. 2530–2540, 2005.

[145] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, "Consistency and stability of tau-leaping schemes for chemical reaction systems," *Multiscale Model. Simul.*, vol. 4, no. 3, pp. 867–895, 2005.

[146] G. Jenkinson and J. Goutsias, "Statistically testing the validity of analytical and computational approximations to the chemical master equation," *J. Chem. Phys.*, vol. 138, pp. 1–19, 2013.

[147] E. W. J. Wallace, D. T. Gillespie, K. R. Sanft, and L. R. Petzold, "Linear noise approximation is valid over limited times for any chemical system that is sufficiently large," *IET Syst. Biol.*, vol. 6, pp. 102–115, 2012.

[148] E. L. Lehman and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New York: Springer, 2005.

[149] P. J. Bickel and K. A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 2007, vol. I.

[150] M. A. Stephens, "Use of the Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables," *J. Roy. Stat. Soc. B Met.*, vol. 32, no. 1, pp. 115–122, 1970.

[151] A. Kolmogoroff, "Confidence limits for an unknown distribution function," *Ann. Math. Statist.*, vol. 12, no. 4, pp. 461–463, 1941.

[152] G. E. Noether, "Note on the Kolmogorov statistic in the discrete case," *Metrika*, vol. 7, no. 1, pp. 115–116, 1963.

[153] H. Ge and H. Qian, "Non-equilibrium phase transition in mesoscopic biochemical systems: from stochastic to nonlinear dynamics and beyond," *J. R. Soc. Interface*, vol. 8, no. 54, pp. 107–116, 2011.

[154] G. Palombo, "Multivariate goodness of fit procedures for unbinned data: An annotated bibliography," *arXiv*, vol. 1102.2407, pp. 1–15, 2011.

[155] C. J. Mecklin and D. J. Mundfrom, "An appraisal and bibliography of tests for multivariate normality," *Int. Stat. Rev.*, vol. 72, pp. 123–138, 2004.

[156] G. Fasano and A. Fanceschini, "A multidimensional version of the Kolmogorov-Smirnov test," *Mon. Not. R. Ast. Soc.*, vol. 255, pp. 155–170, 1987.

[157] R. Beran and M. P. W., "A stochastic minimum distance test for multivariate parameteric models," *Ann. Stat.*, vol. 17, pp. 125–140, 1989.

[158] A. Justel, D. Penña, and R. Zamar, "A multivariate Kolmogorov-Smirnov test of goodness of fit," *Stat. Probabil. Lett.*, vol. 35, pp. 251–259, 1997.

[159] R. D. Cousins, "Annotated bibliography of some papers on combining significances or p-values," *arXiv*, vol. 0705.2209, pp. 1–4, 2007.

[160] W. Feller, "On the Kolmogorov-Smirnov limit theorems for empirical distributions," *Ann. Math. Statist.*, vol. 19, no. 2, pp. 177–189, 1948.

[161] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, 3rd ed. Boca Raton, FL: Chapman & Hall/CRC, 2007.

[162] L. Baringhaus and C. Franz, "On a new multivariate two-sample test," *J. Multivariate Anal.*, vol. 88, pp. 190–206, 2004.

[163] P. R. Rosenbaum, "An exact distribution-free test comparing two multivariate distributions based on adjacency," *J. R. Statist. Soc. B*, vol. 67, no. 4, pp. 515–530, 2005.

[164] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel method for the two-sample problem," *J. Mach. Learn. Res.*, vol. 1, pp. 1–10, 2008.

[165] F. Schlögl, "Chemical reaction models for non-equilibrium phase transitions," *Z. Physik*, vol. 253, pp. 147–161, 1972.

[166] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Statist. Soc. B*, vol. 57, no. 1, pp. 289–300, 1995.

[167] M. Paczuski, S. Maslov, and P. Bak, "Avalanche dynamics in evolution, growth, and depinning models," *Phys. Rev. E*, vol. 53, pp. 414–443, 1996.

[168] J. M. Beggs and D. Plenz, "Neuronal avalanches in neocortical circuits," *J. Neurosci.*, vol. 23, pp. 11 167–11 177, 2003.

[169] L. de Arcangelis and H. J. Herrmann, "Learning as a phenomenon occuring in a critical state," *P. Natl. Acad. Sci. USA*, vol. 107, pp. 3977–3981, 2010.

[170] M. Samoilov, S. Plyasunov, and A. P. Arkin, "Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations," *Proc. Natl. Acad. Sci. USA*, vol. 102, pp. 2310–2315, 2005.

[171] M. Turcotte, J. Garcia-Ojalvo, and G. M. Suel, "A genetic timer through noise-induced stabilization of an unstable state," *P. Natl. Acad. Sci. USA*, vol. 105, pp. 15 732–15 737, 2008.

[172] S. Lapidus, B. Han, and J. Wang, "Intrinsic noise, dissipation cost, and robustness of cellular networks: The underlying energy landscape of MAPK signal transduction," *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 16, pp. 6039–6044, 2008.

[173] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets," *J. Theoret. Biol.*, vol. 22, pp. 437–467, 1969.

[174] ——, *The Origins of Order*. New York, NY: Oxford University Press, 1993.

[175] I. Shmulevich, E. R. Dougherty, K. S., and W. Zhang, "Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks," *Bioinformatics*, vol. 18, pp. 261–274, 2002.

[176] I. Shmulevich, E. R. Dougherty, and W. Zhang, "From Boolean to probabilistic Boolean networks as models of genetic regulatory networks," *Proc. IEEE*, vol. 90, pp. 1778–1792, 2002.

[177] F. Greil and B. Drossel, "Dynamics of critical kauffman networks under asynchronous stochastic update," *Phys. Rev. Lett.*, vol. 95, p. 0487012, 2005.

[178] A. Shreim, A. Berdahl, F. Greil, J. Davidson, and M. Paczuski, "Attractor and basin entropies of random Boolean networks under asynchronous stochastic update," *Phys. Rev. E*, vol. 82, p. 035102(R), 2010.

[179] M. Yang and T. Chu, "Evaluation of attractors and basins of asynchronous random Boolean networks," *Phys. Rev. E*, vol. 85, p. 056105, 2012.

[180] H. Siebert and A. Bockmayr, "Temporal constraints in the logical analysis of regulatory networks," *Theor. Comput. Sci.*, vol. 391, pp. 258–275, 2008.

[181] W. Abou-Jaoudé, D. Ouattara, and M. Kaufman, "From structure to dynamics: Frequency tuning in the p53-Mdm2 network. I. Logical approach," *J. Theor. Biol.*, vol. 258, pp. 561–577, 2009.

[182] G. Vahedi, B. Faryabi, J. Chamberland, A. Datta, and E. Dougherty, "Sampling rate-dependent probabilistic Boolean networks," *J. Theor. Biol.*, vol. 261, pp. 540–547, 2009.

[183] A. Bauer, T. Jackson, Y. Jiang, and T. Rohlf, "Receptor cross-talk in angiogenesis: Mapping environmental cues to cell phenotype using a stochastic, Boolean signaling network model," *J. Theor. Biol.*, vol. 264, pp. 838–846, 2010.

[184] V. Sevim, X. Gong, and J. Socolar, "Reliability of transcriptional cycles and the yeast cell-cycle oscillator," *PLoS Comput. Biol.*, vol. 6, p. e1000842, 2010.

[185] S. Teraguchi, Y. Kumagai, A. Vandenbon, S. Akira, and D. Standley, "Stochastic binary modeling of cells in continuous time as an alternative to biochemical reaction equations," *Phys. Rev. E*, vol. 84, p. 062903, 2011.

[186] G. Stoll, E. Viara, E. Barillot, and L. Calzone, "Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm," *BMC Syst. Biol.*, vol. 6, p. 116, 2012.

[187] R. A. Alberty, *Thermodynamics of Biochemical Reactions*. Hoboken, NJ: Wiley-Interscience, 2003.

[188] W. M. Haddad, V. Chellaboina, and S. G. Nersesov, *Thermodynamics: A Dynamical Systems Approach*. Princeton, NJ: Princeton University Press, 2005.

[189] H. Touchette, "The large deviation approach to statistical mechanics," *Phys. Rep.*, vol. 478, pp. 1–69, 2009.

[190] D. W. Stroock, *An Introduction to Markov Processes*. Berlin: Springer, 2005.

[191] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, "Infectious disease modeling of social contagion in networks," *PLoS Comput. Biol.*, vol. 6:e1000968, no. 11, 2010.

[192] E. M. Broens, C. Espinosa-Gongora, E. A. M. Graat, N. Vendrig, P. J. Van Der Wolf, L. Guardabassi, P. Butaye, J. P. Nielsen, M. C. M. De Jong, and A. W. Van De Giessen, "Longitudinal study on transmission of mrsa cc398 within pig herds," *BMC Vet. Res.*, vol. 8, p. 58, 2012.

[193] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, "Emotions as infectious diseases in a large social network: the sisa model," *Proc. R. Soc. B*, vol. 277, pp. 3827–3835, 2010.

[194] ——, "Infectious disease modeling of social contagion in networks," *PLoS Comput. Biol.*, vol. 6, p. e1000968, 2010.

[195] D. J. Austin and R. M. Anderson, "Studies of antibiotic resistance within the patient, hospitals and the community using simple mathematical models," *Phil. Trans. R. Soc. Lond. B*, vol. 354, pp. 721–738, 1999.

[196] C. Koch and I. Segev, Eds., *Methods in Neural Modeling: From Ions to Networks*, 2nd ed. Cambridge, MA: MIT Press, 1998.

[197] P. C. Bressloff, "Metastable states and quasicycles in a stochastic Wilson-Cowan model of neural population dynamics," *Phys. Rev. E*, vol. 82:051903, 2010.

[198] M. Diaz, *Petri Nets: Fundamental Models, Verification and Applications*. Hoboken, New Jersey: Wiley-ISTE, 2009.

[199] G. C. M. A. Ehrhardt, M. Marsili, and F. Vega-Redondo, "Phenomenological models of socioeconomic network dynamics," *Phys. Rev. E*, vol. 74:036106, 2006.

[200] T. Gross, C. J. D. D'Lima, and B. Blasius, "Epidemic dynamics on an adaptive network," *Phys. Rev. Lett.*, vol. 96:208701, 2006.

[201] P. Holme and M. E. J. Newman, "Nonequilibrium phase transition in the coevolution of networks and opinions," *Phys. Rev. E*, vol. 74:056108, 2006.

[202] C. Zhou and J. Kurths, "Dynamical weights and enhanced synchronization in adaptive complex networks," *Phys. Rev. Lett.*, vol. 96:164102, 2006.

[203] T. Gross and B. Blasius, "Adaptive coevolutionary networks: a review," *J. R. Soc. Interface*, vol. 5, pp. 259–271, 2008.

[204] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, "Noise bridges dynamical correlation and topology in coupled oscillator networks," *Phys. Rev. Lett.*, vol. 104:058701, 2010.

[205] J. Zhang, C. Zhou, X. Xu, and M. Small, "Mapping from structure to dynamics: A unified view of dynamical processes on networks," *Phys. Rev. E*, vol. 82:026116, 2010.

[206] W.-J. Yuan and C. Zhou, "Interplay between structure and dynamics in adaptive complex networks: Emergence and amplification of modularity by adaptive dynamics," *Phys. Rev. E*, vol. 84:016116, 2011.

[207] R. Heinrich and S. Schuster, *The Regulation of Cellular Systems.* New York: Chapman & Hall, 1996.

[208] A. Varma, M. Morbidelli, and H. Wu, *Parametric Sensitivity in Chemical Systems.* Cambridge: Cambridge University Press, 1999.

[209] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, "Sensitivity analysis for chemical models," *Chem. Rev.*, vol. 105, pp. 2811–2827, 2005.

[210] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis: The Primer*. New York: Wiley, 2008.

[211] H.-X. Zhang, W. P. Dempsey Jr, and J. Goutsias, "Probabilistic sensitivity analysis of biochemical reaction systems," *J. Chem. Phys.*, vol. 131:094101, 2009.

[212] H.-X. Zhang and J. Goutsias, "A comparison of approximation techniques for variance-based sensitivity analysis of biochemical reaction systems," *BMC Bioinf.*, vol. 11:246, 2010.

[213] ——, "Reducing experimental variability in variance-based sensitivity analysis of biochemical reaction systems," *J. Chem. Phys.*, vol. 134:114105, 2011.

[214] D. Kim, B. J. Debusschere, and H. N. Najm, "Spectral methods for parametric sensitivity in stochastic dynamical systems," *Biophys. J.*, vol. 92, pp. 379–393, 2007.

[215] S. Plyasunov and A. P. Arkin, "Efficient stochastic sensitivity analysis of discrete event systems," *J. Comput. Phys.*, vol. 221, pp. 724–738, 2007.

[216] S. H. Dandach and M. Khammash, "Analysis of stochastic strategies in bacterial competence: A master equation approach," *PLoS Comput. Biol.*, vol. 6:e1000985, no. 11, 2010.

[217] K. H. Kim and H. M. Sauro, "Sensitivity summation theorems for stochastic biochemical reaction systems," *Math. Biosci.*, vol. 226, pp. 109–119, 2010.

[218] M. Rathinam, P. W. Sheppard, and M. Khammash, "Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks," *J. Chem. Phys.*, vol. 132:034103, 2010.

[219] M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf, "Sensitivity, robustness, and identifiability in stochastic chemical kinetics models," *Proc. Natl. Acad. Sci. USA*, vol. 108, no. 21, pp. 8645–8650, 2011.

[220] D. F. Anderson, "An efficient finite difference method for parameter sensitivities of continuous time Markov chains," *SIAM J. Numer. Anal.*, vol. 50, pp. 2237–2258, 2012.

[221] P. B. Warren and R. J. Allen, "Steady-state parameter sensitivity in stochastic modeling via trajectory reweighting," *J. Chem. Phys.*, vol. 136:104106, 2012.

[222] E. S. Wolf and D. F. Anderson, "A finite difference method for estimating second order parameter sensitivities of discrete stochastic chemical reaction networks," *J. Chem. Phys.*, vol. 137, p. 224112, 2012.

[223] R. Srivastava, D. F. Anderson, and J. B. Rawlings, "Comparison of finite difference based methods to obtain sensitivities of stochastic chemical kinetic models," *J. Chem. Phys.*, vol. 138, p. 074110, 2013.

[224] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: A comparison of global optimization methods," *Genome Res.*, vol. 13, pp. 2467–2474, 2003.

[225] E. J. Crampin, S. Schnell, and P. E. McSharry, "Mathematical and computational techniques to deduce complex biochemical reaction mechanisms," *Prog. Biophys. Mol. Bio.*, vol. 86, pp. 77–112, 2004.

[226] G. Maria, "A review of algorithms and trends in kinetic model identification for chemical and biochemical systems," *Chem. Biochem. Eng. Q.*, vol. 18, no. 3, pp. 195–222, 2004.

[227] A. Golightly and D. J. Wilkinson, "Bayesian inference for stochastic kinetic models using a diffusion approximation," *Biometrics*, vol. 61, pp. 781–788, 2005.

[228] ——, "Bayesian sequential inference for stochastic kinetic biochemical network models," *J. Comput. Biol.*, vol. 13, no. 3, pp. 838–851, 2006.

[229] S. Reinker, R. M. Altman, and J. Timmer, "Parameter estimation in stochastic biochemical reactions," *IEE Proc. - Syst. Biol.*, vol. 153, no. 4, pp. 168–178, 2006.

[230] R. J. Boys, D. J. Wilkinson, and T. B. L. Kirkwood, "Bayesian inference for a discretely observed stochastic kinetic model," *Stat. Comput.*, vol. 18, pp. 125–135, 2008.

[231] M. Komorowski, B. Finkenstädt, C. V. Harper, and D. A. Rand, "Bayesian inference of biochemical kinetic parameters using the linear noise approximation," *BMC Bioinf.*, vol. 10:343, 2009.

[232] S. K. Poovathingal and R. Gunawan, "Global parameter estimation methods for stochastic biochemical systems," *BMC Bioinf.*, vol. 11:414, 2010.

[233] Y. Wang, S. Christley, E. Mjolsness, and X. Xie, "Parameter inference for discretely observed stochastic kinetic models using stochastic gradient descent," *BMC Syst. Biol.*, vol. 4:99, 2010.

[234] B. J. Daigle Jr., M. K. Roh, L. R. Petzold, and J. Niemi, "Accelerated maximum likelihood parameter estimation for stochastic biochemical systems," *BMC Bioinf.*, vol. 13:68, 2012.

[235] J. C. Spall, *Introduction to Stochastic Search and Optimization.* Hoboken, New Jersey: John Wiley & Sons, 2003.

[236] D. Colquhoun, K. A. Dowsland, M. Beato, and A. J. Plested, "How to impose microscopic reversibility in complex reaction mechanisms," *Biophys. J.*, vol. 86, pp. 3510–3518, 2004.

[237] W. Liebermeister and E. Klipp, "Bringing metabolic networks to life: Convenience rate law and thermodynamic constraints," *Theor. Biol. Med. Model.*, vol. 3:41, 2006.

[238] J. Yang, W. J. Bruno, W. S. Hlavacek, and J. E. Pearson, "On imposing detailed balance in complex reaction mechanisms," *Biophys. J.*, vol. 91, pp. 1136–1141, 2006.

[239] G. Jenkinson, X. Zhong, and J. Goutsias, "Thermodynamically consistent Bayesian analysis of closed biochemical reaction systems," *BMC Bioinf.*, vol. 11:547, 2010.

[240] G. Jenkinson and J. Goutsias, "Thermodynamically consistent model calibration in chemical kinetics," *BMC Syst. Biol.*, vol. 5:64, 2011.

[241] W. G. Jenkinson and J. Goutsias, "A screening method for dimensionality reduction in biochemical reaction system calibration," in *Proceedings of the 2010 IEEE International Workshop on Genomic Signal Processing and Statistics*, Cold Spring Harbor Laboratory, New York, November 10-12 2010, pp. 214–219.

[242] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge University Press, 1985.

[243] ——, *Topics in Matrix Analysis*. New York, NY: Cambridge University Press, 1991.

[244] L. Billard, "Factorial moments and probabilities for the general stochastic epidemic," *J. Appl. Prob.*, vol. 10, pp. 277–288, 1973.

Vita

W. Garrett Jenkinson was born on May 1, 1985 in Stuart, FL. He received the B.S. and M.S. degrees in Electrical and Computer Engineering (ECE) from Carnegie Mellon University in 2007, and enrolled in the ECE Ph.D. program at The Johns Hopkins University later that year. Garrett was awarded the M.S.E. degree in Applied Mathematics and Statistics from The Johns Hopkins University in 2009, as well as the C.A.S. degree in Nanobiotechnology from The Johns Hopkins University in 2013.

At Carnegie Mellon, Garrett was inducted into the Eta Kappa Nu Honors Society; graduated with University, College, and Department Honors; and was awarded the Lockheed Martin Award for Best Undergraduate ECE Research, the David Tuma Award for Most Outstanding Undergraduate Laboratory Work, and Second Place Overall in CIT's Undergraduate Engineering Research Competition. At Johns Hopkins, he received the Abel Wolman Fellowship, the NSF IGERT Fellowship, the DoD NDSEG Fellowship, and was named a Siebel Scholar for the class of 2013. Garrett co-lectured and co-developed "Principles of Complex Networked Systems", a graduate

course that is cross-listed in the Department of Electrical and Computer Engineering and the Department of Biomedical Engineering at Johns Hopkins.

Starting in October 2013, Garrett will become a post-doctoral fellow at The Johns Hopkins University School of Medicine in the Complex Systems Science Laboratory as well as the Center for Epigenetics, where he will be studying the role of stochastic epigenetic variation on evolution and complex human diseases, such as cancer.